Protecting people from illegal harms online

Register of Risks

Superseded

For the version in force see: Online safety regulatory documents and guidance

Published 16 December 2024

Contents

Section

Introduction to the causes and impacts of illegal harm online (Register of Risks) 4

1. Terrorism 32
2. Child Sexual Exploitation and Abuse (CSEA) 52

2A. Grooming 64

2B. Child Sexual Abuse Material (CSAM) 83

3. Hate 103
4. Harassment, stalking, threats and abuse 121
5. Controlling or coercive behaviour (CCB) 143
6. Intimate image abuse 163
7. Extreme pornography offence 185
8. Sexual exploitation of adults 196
9. Human trafficking 207
10. Unlawful immigration 225
11. Fraud and financial services offences 234
12. Proceeds of Crime 255
13. Drugs and psychoactive substances 265
14. Firearms, knives and other weapons 281
15. Encouraging or assisting suicide (or attempted suicide) 291
16. Foreign interference offence 310
17. Animal cruelty 334

18. Epilepsy trolling offence 349
19. Cyberflashing 357
20. Encouraging or assisting serious self-harm 365
21. False communications 384
22. Obscene content showing torture of humans and animals (the s.127(1) offence) 392
23. Threatening communications 401
24. Search services 403
25. Governance, systems and processes 416

Annex

A1. Glossary of terms 434
A2. Updating the Register of Risks 446
A3. Updating the Risk Profiles 458

Introduction to the causes and impacts of illegal harm online (Register of Risks)

The **Illegal Harms Register of Risks** ('Register of Risks') is our assessment of the causes and impacts of illegal online harms based on the evidence that we have gathered over the past four years. The Register of Risks presents our full risk assessment of where and how illegal harms manifest online and the characteristics of services that are relevant to the risks of harm. It forms part of our duty under the Act to assess the factors that can cause a risk of harm to individuals on a service.

Ensuring service providers undertake a high-quality risk assessment is one of our strategic objectives and the Register of Risks is an important resource for achieving this. It is intended to act as a central resource for service providers when they are conducting their risk assessments, providing a clear understanding of how harms manifest online and how specific characteristics of services and users play a role. The risk factors identified in the Register of Risks also inform the **Risk Profiles** that sit within the Risk Assessment Guidance, helping service providers identify the areas of greatest likely risk due to particular characteristics in the design, functionality and user base of their services. In their risk assessment, they will be expected to assess the likelihood and impact of those risks.

Our assessment focuses on the over 130 priority offences defined in the Act. For ease of navigation and due to similarities in how some kinds of illegal content manifest, we have grouped these into 17 broad kinds of illegal harm. These include illegal harms such as child sexual exploitation and abuse (CSEA), terrorism, fraud, hate speech, weapons and drugs offences, modern slavery and human trafficking, foreign interference and most recently, animal cruelty. We summarise the findings of this assessment below and set out the detailed analysis in the subsequent Register of Risks chapters – one for each kind of illegal harm. We also have six additional chapters, each covering a relevant non-priority offence, such as encouraging or assisting serious self-harm, false communications and 'epilepsy trolling'.

Based on the evidence contained in the illegal harms Register of Risks, we've identified risk factors for each kind of harm – the characteristics of online services and users that we believe increase the likelihood of illegal content being encountered, and the associated risk of harm.

Offenders often rely on different types of services to commit or facilitate the offences covered by the Act. For instance, both fraudsters and perpetrators of grooming will often contact potential victims on public forums and then seek to move them onto private, sometimes encrypted, messaging services. This means that action to tackle online harms cannot focus exclusively on a small subset of services and cannot be targeted exclusively at the largest services. Rather, it needs to address a broad range of service types including both large services and the many smaller services in scope of the Act.

The role of the new online safety regulations is to get services to put in place safeguards which allow users to enjoy the benefits they bring while managing the risks appropriately.

Online harms and the factors which can increase the risk of harm occurring are changing all the time as technology develops. The recent emergence of generative AI provides a particularly clear example of this. As well as bringing important benefits, generative AI creates new risks. Image-generation models, for example, can be used in some cases to create child sexual abuse material (CSAM). Studies have also highlighted the use of generative AI to create 'deepfakes' in support of foreign interference campaigns. They have also been used to generate instructions for how to access unlicensed firearms and to create authentic sounding audio or textual messages to deceive victims in cases of fraud.

The constant emergence of new risks makes it important that services conduct regular risk assessments. It also makes robust corporate governance particularly important. Where services have good governance arrangements in place with clear accountability for managing risks, they are more likely to detect and appropriately manage emerging risks. In addition to recommending measures to address specific harms, a key focus for us is ensuring service providers conduct robust risk assessments and have appropriate governance arrangements in place.

About this document

This introduction to the Register of Risks is structured as follows:

a) **Section 1: Introduction to the causes and impacts of illegal harm online (*this* section):** provides a summary of our findings and an introduction to how illegal harm manifests online. We set out some concepts and context useful for interpreting the huge range of issues covered in the Register of Risks itself (see Section 3)

b) **Section 2: Ofcom's Register of Risks for illegal harms:**

   i) **Methodology for conducting our risk assessment:** this chapter provides an overview of the methodology used to conduct our sector-wider risk assessment, and introduces concepts presented in the Register of Risks

   ii) **Register of Risks:** This is our sector-wide risk assessment. It identifies and assesses the risk of physical and psychological harm to individuals in the UK presented by regulated user-to-user (U2U) and search services and identifies the characteristics of services relevant to such risks of harm. Our analysis is presented in X chapters, one for each kind of illegal harm.

Summary of Findings

Over the past three years we have conducted an extensive analysis of the causes and impacts of illegal online harms. As part of our analysis, we reviewed thousands of sources from hundreds of research organisations, academic institutions, online service providers, government, law enforcement and civil society organisations.

In the November 2023 Consultation we published an initial draft of this analysis. This drew on research that we have commissioned, a comprehensive review of the existing published evidence and engagement with a wide variety of stakeholders. The 199 responses to the November 2023 consultation contained a large body of additional evidence on and insights into the illegal harms in scope of the Act. Over the past year, we have analysed these responses in detail and conducted follow up research, analysis and stakeholder engagement to deepen our understanding of the harms.

Taken together, the work we have done over the past three years shows that illegal online content is widespread and, in many cases, growing in prevalence. For example, Ofcom's own research found that 87% of adult internet users report having encountered a scam or fraud online and 25% of these people have lost money as a result.[1] Almost a fifth of children experienced sexual solicitation from adults they have chatted with online. In a recent report the NSPCC stated more than 7,000 Sexual Communication with a Child offences were recorded by Police in 2023/24, an 89% increase since this offence came into force in 2017/18.[2]

Given the breadth of the risks online, anyone can experience harm in some capacity, just like anyone can be a victim of crime offline – 68% of UK internet users reported having encountered potentially harmful content online in the past four weeks.[3] But in most instances, the risks people face online are not equal, with children and people with certain protected characteristics most likely to be affected and certain kinds of harm being significantly more prevalent among certain groups. Anyone can fall prey to the right kind of online fraud; but it is not a surprise that victims of various forms of online harassment are so often women, mirroring and potentially even amplifying wider challenges in society. Studies have shown that women are five times more likely to be victims of intimate image abuse. 30% of minority ethnic internet users report having encountered 'hateful, offensive or discriminatory content', compared to 25% of all internet users.[4] Generally, the more protected characteristics or vulnerabilities someone has, the greater the risk of harm they face from priority illegal harms in the Act.

Therefore, since the impact of the harms we have looked at can be extremely severe, our work has never been more critical than it is today. Harm is not limited to the online world and profoundly affects people's lives. Our updated Register of Risks demonstrates that harms, such as online grooming, can cause lifelong negative psychological impacts for victims. Additionally, the harm experienced may not be limited to the individuals who directly encounter risks or illegal content online. It also impacts those who are the victims of the subsequent actions of those people who have encountered illegal content or participated in illegal activity online. In some cases – in relation to terrorism, for example, or the erosion of trust in democratic processes caused by state-sponsored disinformation campaigns – it can be argued that these harms have a wider societal impact.

We have seen how all types of services can pose a risk of harm from the priority illegal content addressed by the Act, individually and when used together to facilitate criminal activity. While many

[1] Ofcom, 2023. Online Scams & Fraud Research. [accessed 21 November 2024].
[2] NSPCC, 2024. Online grooming crimes against children increase by 89% in six years. [accessed 14 November 2024].
[3] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 28 November 2024].
[4] Ofcom, 2024.

service providers have made significant investments in tackling online harms in recent years, our evidence shows these have not yet been sufficient. For instance, encrypted cloud-based file-sharing services are still used by offenders to store and share CSAM; social media services with large numbers of targetable users and with highly effective content recommender systems remain effective places for state-sponsored disinformation campaigns to take place. Our analysis and engagement with stakeholders also shows how the risks to the UK public evolve in step with the online landscape and are arguably more significant now than they ever have been.

Our work in the past year shows that the kinds of illegal harm we have looked at occur on services of all types. Services as diverse as social media services, dating services, marketplaces and listings services, search services, user-to-user pornography services, and file-storage and file-sharing services are all used to disseminate some of the types of illegal content and facilitate the offences we have looked at in the Register of Risks. While certain characteristics of a service might make a certain type of harm more likely to occur on or via a particular service, no service in-scope of the Act is entirely immune from risk simply due to the nature of its design, user base or the way the service provider conducts business and runs their organisation.

For instance, the size of a service does not dictate the level of risk but influences how and what kinds of illegal harm are more likely to manifest. Bad actors use both large and small services to spread illegal content, though the way in which they use these services is likely to differ. For example, terrorists often use large services to disseminate propaganda to large audiences, but often use small services for more covert activities such as recruitment, planning and fundraising.

Although a very wide range of service types pose risks of the priority illegal harms in the Act, there are certain service types that appear to play a particularly prominent role in the spread of priority illegal content. For example, our updated analysis suggests that video-sharing services and dating services pose a particular high risk with regards to hate offences and drugs and psychoactive substances offences respectively.

Similarly, certain 'functionalities' stand out as posing particular risks because of the prominent role they appear to play in the spread of illegal content and the commission and facilitation of offences:

- **End-to-end encryption**: Offenders often use end-to-end encrypted services to evade detection. For example, end-to-end encryption can enable perpetrators to circulate CSAM, engage in fraud, and spread terrorist content with a reduced risk of detection.
- **Pseudonymity and anonymity**: In most cases where offenders are using online services to engage in illegal activity, hiding their identity is incredibly important as it supports their ability to evade detection. There is also some evidence that pseudonymity (where a person's identity is hidden from others through the use of aliases) and anonymity can embolden offenders to engage in a number of harmful behaviours with reduced fear of consequences who otherwise might not have. For example, while the evidence is contested, some studies suggest that pseudonymity and anonymity can embolden people to commit hate speech. At the same time, cases of harassment and stalking often involve perpetrators creating multiple fake user profiles to contact individuals against their will and to circumvent blocking and moderation.
- **Livestreaming**: There are many examples of terrorists livestreaming attacks, this can in turn incite further violence. The use of livestreaming remains a persistent feature of far-right lone attackers, many of whom directly reference and copy aspects of previous attacks. Similarly, perpetrators can exploit livestreaming functionality when abusing children online.

- **Content recommender systems**: Content recommender systems are commonly designed to optimise for user engagement and learn about users' preferences. Where a user is engaging with harmful content such as hate speech or content which promotes suicide, there is a risk that this might result in ever more of this content being served up to them.

Online harms and the risk factors which cause them are changing all the time as technology develops and society evolves. The recent emergence of generative AI provides a particularly clear example of this. As well as bringing important benefits, generative AI creates new risks across a variety of kinds of illegal harm including CSAM, terrorism, fraud and foreign interference.

The functionalities we describe above are not inherently harmful and can have important benefits for users. End-to-end encryption plays an important role in safeguarding privacy online. Pseudonymity and anonymity can allow people to express themselves and engage freely online. In particular, anonymity can be important for historically marginalised groups such as members of the LGBTQ+ community who wish to talk openly about their sexuality or explore gender identity without fear of discrimination or harassment. Recommender systems benefit internet users by helping them find content which is interesting and relevant to them.

The constant emergence of new risks makes it vital that services conduct regular risk assessments. It also makes robust corporate governance particularly important. Where services have good governance arrangements in place with clear accountability for managing risks, they are more likely to detect and appropriately manage emerging risks. In addition to recommending measures to address specific harms, a key focus for us as the Online Safety regime comes into force will, therefore, be ensuring that services conduct robust risk assessments and have appropriate governance arrangements in place.

As we explain in 'Our approach to developing Codes measures', we have designed measures in our Codes of Practice to target high-risk service types and functionalities. The role of the new online safety regulations is not to restrict or prohibit the use of such functionalities, but rather to get service providers to put in place safeguards which allow users to enjoy the benefits they bring while managing the risks appropriately.

Understanding illegal harms online

Most of our analysis of the risks of harm is focused on specifics: every chapter of the Register of Risks is focused on one kind of illegal harm, in some cases individual offences, and seeks to present a non-exhaustive set of risk factors that we have identified from the evidence. We recognise that every instance of harm experienced or facilitated online is also unique, and risk is a multi-faceted, often highly complex and, at times, personal issue.

We have identified some broad trends that underpin how harm can manifest online that can be seen in many, if not all, instances where illegal activity occurs or is facilitated by online services. Understanding these trends is helpful when interpreting the huge variety of risks and harms explored in the Register of Risks.

- **Perpetrator behaviour is diverse and adapts to the online environment:** In most instances, it is intentional misuse of online services that leads to illegal harm, whether directly or indirectly. As a term "perpetrators" – those who commit offences online – describes a very broad range of people and organisations. For example, perpetrators can be state-sponsored agencies or form part of organised crime groups or networks. In other cases, a perpetrator may be an individual operating alone to commit an offence. The circumstances that lead to the use of online services to commit illegal acts are also multi-faceted and complex. The mitigations service providers put in place must therefore be able to disrupt a huge variety of intentional criminal activity – from using messaging services to advertise illegal knives and drugs for sale to offenders sharing 'how to' guides to groom children online – while also preventing other users from encountering illegal content.
- **Functionalities, features and technologies:** Any functionality, feature or technology which facilitates the spread of information and makes it easier for people to interact with one another necessarily facilitates the spread of both good and bad content and enables positive and harmful interactions. The functions that deliver value to users in some contexts can and are misused by perpetrators to achieve their own aims. For instance, predators contact children using the same direct messaging functionality designed for children to keep in touch with their friends and family, hiding their real identity with fake user profiles. In some cases, functionalities simply do what they are designed to do, but with limited regard to the negative consequences resulting from working as intended – a system designed to maximise engagement or provide accurate responses to queries has no reason not use potentially illegal or harmful content to do so unless designed not to.
- **Services:** Online services are where these functionalities and the people who use and misuse them come together. They provide the specific combinations of functionality and access to people and content, along with the systems, processes, governance and even culture, that make them uniquely suited to delivering huge value for users, but also the wide variety of harm. For example, encrypted cloud-based file-sharing services are used by offenders to store and share CSAM; social media services with large numbers of targetable users and with highly effective content recommender systems are effective places for state-sponsored disinformation campaigns to take place.
- **Victim's experiences:** Given the breadth of illegal harm that manifests online everyone can be at risk of some form of online harm. But in most instances, there are certain people who are more vulnerable, more susceptible, and more likely to experience severe harm as a result of their experience. Anyone can fall prey to the right kind of

online fraud; but the victims of various forms of online harassment are disproportionately women.

Every instance of online harm requires some combination of these components. By better understanding the specific instances in which they come together to cause harm – as we have set out in our Register of Risks – we can better minimise the risks and prevent harm from occurring.

Perpetrator behaviours

As a term "perpetrators" – those who commit offences online – describes a very broad range of people and organisations, acting alone or as part of collective efforts. For example, perpetrators can be state-sponsored agencies or form part of organised crime groups or networks. In other cases, the perpetrator may be an individual operating alone to commit the offence. There are also cases of preparators supporting one another to provide advice or share content with each other among an informal network of offenders. For example, child sexual abuse material (CSAM) is spread through offender-to-offender networks, as well as perpetrators sharing advice and tips with each other in user groups.

Service providers should be mindful that perpetrators also 'service-hop', moving between different services in the course of carrying out certain offences. Perpetrators may meet their victims and survivors, or other perpetrators, on one service and then move their interactions to another service. Moving between services is often driven by the unique benefits each service affords to this activity, based on the characteristics of the service. For example, in a case of grooming, a perpetrator may choose to identify a child and initiate contact with them on a social media service that is well-suited to finding and initiating contact with other (child) users, and then move their communication to a private, likely encrypted messaging service where less moderation is expected. The journey of an offence, and the point at which a service is most likely to be used by a potential perpetrator will be relevant to a service provider's risk assessment. As the accuracy and consistency of moderation efforts continue to improve and other measures restrict the opportunities afforded to perpetrators, particularly on the largest most commonly used services, we expect to see more of this activity displaced to online spaces perpetrators perceive to be more privacy-preserving.

The channel of communication used to share content and interact with other users is particularly important and perpetrators use both open and closed channels to commit offences. Open channels which allow for more visible one-to-many communication, such as posting on social media groups, are more likely to be used when perpetrators want to maximise the number of people they are disseminating content to, or want to fish for potential victims in the widest possible pool.[5] Conversely, perpetrators are more likely to use closed channels, such as end-to-end encrypted messaging services, for communication they do not want to be visible to others or to the service providers. This might be discussing illegal activity or planning offences covertly or having private interactions with victims and survivors.[6]

[5] This could be the case for fraud, hateful content or promotion of suicide content, for example.

[6] For example, perpetrators often message children on private messaging services in grooming offences. Similarly, perpetrators have been known to share CSAM with one another on private messaging services.

Service functionalities, features and technologies

Any functionalities which facilitate communication or dissemination of content create opportunities and risks

The functionalities of online services, in general, are not inherently positive nor negative. They facilitate communication at scale and reduce friction in user-to-user interactions, making it possible to disseminate both positive and harmful content. For example, users can engage with one another through direct messaging and livestreaming, develop relationships and reduce social isolation. However, these same functionalities can also enable the sharing of illegal material such as livestreams of terrorist atrocities or messages sent with the intent of grooming children.

Many functionalities are common across a wide range of services, and therefore a very large number of services can potentially pose some risks of harm to individuals. Whether functionalities ultimately create opportunities for positive engagement or lead to risks of harm will depend on the service's governance, systems and processes put in place to mitigate risk.

Generative artificial intelligence (GenAI)

GenAI delivers a range of benefits. For example, it powers a range of features and functionalities online, from summarising search results to creating avatars or new gaming environments. The underlying models and tools also can be utilised by service providers to mitigate risk online, improving the efficiency, accuracy and scale of tools that improve user safety.

However, there is emerging evidence that GenAI technologies can be used to facilitate or commit harm on user-to-user (U2U) and search services, although given the rapid developments in the technology, it can be difficult to quantify the risk of harm to individuals and new evidence of risk of harm is likely to emerge in the future.

In particular, AI-generated child sexual abuse material (CSAM) has been identified as a significant and growing problem[7], with the National Crime Agency (NCA) warning of the ease and availability of AI-generated CSAM contributing to a "*normalisation of offending behaviour*" and creating "*a more permissive environment for perpetrators*" which will put children at increased risk.[8] Reports of deepfake intimate image abuse in the past year have also increased significantly[9], alongside a dramatic increase in the volume of services that 'nudify' targets, enabling users to create non-consensual intimate images of anyone they have images of.

The use of GenAI to generate realistic content intended to deceive, and to do this with ease or at scale is also particularly important. Deepfake content is being used increasingly in cases of fraud, with highly targeted deepfakes used to defraud individuals or designed to deceive a mass audience of thousands or even millions.[10] UK Finance has highlighted specific risks GenAI technologies pose in regard to fraud, including image generation to produce images of real or fictional people for the purpose of fraud; deepfake audio which could be used to impersonate a victim's contact to persuade

[7] See the Register of Risks chapter Child Sexual Abuse and Exploitation for further information

[8] National Crime Agency, 2024. Technological Tipping Point Reached in Fight Against Child Sexual Abuse. [accessed 14 October 2024]

[9] In 2023, more deepfake abuse videos were posted online than in every previous year combined

[10] For example, a well-known fake advert featuring a likeness of Martin Lewis was shared on Facebook, in which he appeared to be asking users to sign up for a non-existent Elon Musk investment. Deepfakes can also be used in romance scams, with fraudsters using GenAI and related tools to create inauthentic profiles with images of non-existent people. More elaborate romance scams have seen fraudsters take part in live deepfake video calls with their victims. Sources: BBC, 2023. Martin Lewis felt 'sick' seeing deepfake scam ad on Facebook. [accessed 14 October 2024]; Burgess, M. 2024. The Real-Time Deepfake Romance Scams Have Arrived, Wired, 14 April. [accessed 14 October 2024]

them to make a payment, or potentially to pass voice identity verification; and AI text generation to make it easier to produce fraudulent messages at greater speed and volume.[11]

GenAI technology can be used in foreign interference campaigns. Across the world, there has been evidence of the use of AI-generated audio, image, video and text-based content, by both state-linked and non-state actors, to influence elections and public opinion more broadly.[12] This can include creating false or misleading videos of state figures.

Recent reports also suggest that GenAI chatbots[13] have been shown to offer advice that promotes eating disorders[14] and encourages self-harm.[15] There are also reports of early GenAI chatbots generating instructions for how to access unlicensed firearms and how to make explosive materials as well as dangerous chemicals[16] which could be used to cause harm.[17]

The Online Safety Act is technology neutral and AI-generated content which is illegal and shared on a U2U service or present in search results needs to be treated the same way as all other illegal content.[18]

Recommender systems can deliver benefits but can also increase the risks of harm

Many services use recommender systems, which deliver a range of benefits to users. They personalise each user's experience by helping them find content they are likely to want to engage with and should be considered in the context of each specific service, the kind of content accessible via that service and the way the recommender system is designed. However, they can also increase risk.

Our evidence has found that two types of recommender systems could potentially increase risk: content recommender systems and network recommender systems.

Content recommenders

Content recommenders are algorithmic systems which determine the relative ranking of an identified pool of content, including user-generated content from multiple users, on content feeds. These systems may amplify risks of harm in a number of ways. Firstly, they can push users into 'filter bubbles', where users will only see content similar to other content they engage with and there are limited, or no, opportunities to be exposed to opposing views. In regard to illegal harms, risk stems from the chance that once a user has been exposed to illegal content or engaged with it, they will

---

[11] UK Finance, 2024. The impact of AI in financial services: opportunities, risks and policy considerations. [accessed 18 November 2024].

[12] See the Register of Risks chapter Foreign interference for a detailed discussion of the implications of GenAI in this area.

[13] An AI chatbot is an automated software program that uses artificial intelligence and natural language processing to simulate a conversation.

[14] The US-based National Eating Disorder Association had taken down its chatbot, "Tessa", after reports that it had been providing inappropriate content, including advice on how to lose weight. Source: Aratani, L, 2023. US eating disorder helpline takes down AI chatbot over harmful advice, *The Guardian,* 31 May. [accessed 5 October 2023].

[15] During OpenAI's 'red-team' exercises, a term describing the systematic use of adversarial testing methods to probe and address the risk of language models producing harmful outputs, researchers found that early versions of the company's new GPT-4 generative AI model was capable of generating advice or encouragement for self-harm behaviours, for example self-mutilation. Source: OpenAI, 2023. gpt-4-system-card. [accessed 5 October 2023].

[16] Rose, J, 2022. OpenAI's New Chatbot Will Tell You How to Shoplift And Make Explosives, *Vice,* 1 December. [accessed 5 October 2023]; OpenAI, 2023. gpt-4-system-card. [accessed 5 October 2023].

[17] Given the limited evidence of this emerging technology, we have not included some of the examples as part of the detailed risk analysis presented in the Register. We will continue to monitor the landscape with the expectation that more evidence showing a risk of harm associated with GenAI will emerge.

[18] For more information on GenAI in the Online Safety Act (2023) see Ofcom's Open letter to UK online service providers regarding Generative AI and chatbots

continue to be pushed more and more similar content.[19] Users may also be pushed into 'rabbit holes'. This describes a situation where the user is encouraged (by algorithms) to consume more and more content on the same theme, and which can become more extreme or harmful over time, potentially leading to illegal content.[20]

Potential perpetrators can also 'game' recommender systems, to cause an outcome favourable to them as a result of abusing the algorithm by, for example, continuously posting or pushing extreme content.[21] [22] [23] By design, content recommender systems also serve content to users they have not actively chosen to see, increasing the risk that any one individual encounters content that is harmful to them.

The relationship between specific kinds of illegal or otherwise harmful content and recommender systems is set out in the harms-specific chapters in this volume. For several kinds of illegal harm, content recommender systems emerge as an important risk factor in how relevant content spreads and is accessed online.

Recommender systems are also often an integral part of a service provider's business model as they form part of a strategy for maximising engagement. Through lengthening the time spent or level of active engagement with content, service providers can increase revenue generated through advertising, for example, by serving more, or more targeted, adverts to users. The risks associated with advertising-based business models originate primarily from recommender system design. See 'A service's business model and commercial profile can increase risk' further in this chapter for more detail.

Product recommenders

Product recommenders are technically distinct from content recommender systems, but the two technologies are not legally distinguishable due to the OSA definition of 'content' being inclusive of product listings. We distinguish between 'product' and 'content' recommender systems for two reasons. First, there is a lack of evidence showing product recommender systems contributing to the dissemination of illegal content. Second, they are a distinct technology which would require a separate policy approach from content recommender systems.

We also do not have evidence suggesting product recommender systems directly recommend user-generated attributes of product listings, or affect their visibility, unlike the way relevant content recommender systems do. Available research leads us to believe that although product listings may include user-generated attributes (for example, images, text, and videos), users primarily interact with (click, view, purchase) non-user-generated attributes such as product categories, price range, and brands. The evidence we have suggests product recommender systems use algorithms to

[19] The evidence available suggests this could be a risk for CSAM, hateful content and suicide and self-harm.
[20] Ofcom-commissioned research has also shown that design choices can influence the extent to which users are led on 'pathways' from benign to increasingly harmful content (also known as 'rabbit holes'). Studies have shown that these effects can increase user exposure to a number of harmful content types, including self-harm content, eating disorder content and extreme. However, design choices are not the only factor shaping user exposure to illegal content. Some researchers, for example, have argued that rabbit-hole effects occur more frequently in cases where users are already inclined to seek out harmful content.
[21] Gaming of the algorithms can be achieved by using keyword and tagging features or by coordinating a large number of queries or posts to inflate the popularity of a query or post. Because recommender systems can deliver personalised content in an engaging manner, they can be an attractive target for extremist movements and might also be used to draw users towards more niche, ephemeral, or anonymised social media services which are harder to moderate.
[22] Design choices can also determine the likelihood of recommender systems being gamed by bad actors. This can happen if the design of a system is too simplistic (e.g. with only a small number of information signals feeding into its ranking decisions) or if granular details of the design are made publicly available.
[23] Evidence suggests that this is relevant for the foreign interference offence, and harassment, stalking and threat offences.

analyse only user interactions with the non-user-generated attributes of content to make product recommendations.[24]

Network recommenders

Network recommenders are a type of recommender system that suggests users or groups of users to connect with. Network recommenders may consider a variety of user interactions, mutual connections, and group memberships to determine which network recommendations might be relevant and useful. These systems may amplify risks of harm by helping offenders to find each other by connecting like-minded individuals, or by helping offenders to find potential victims. This is particularly relevant in grooming, as offenders who attempt to connect with multiple child users can then be recommended to other child users.

Online services

A range of services can attract perpetrators

Our analysis of U2U services found that some service types are used to facilitate and commit a wide range of offences. Social media services and private messaging services, in particular, were found to pose risks of several different kinds of illegal harm, including fraud, terrorism, CSAM and the foreign interference offence. This may be due to their popularity and the functionalities that are typically found on them.

Other service types can be used in more targeted ways to facilitate and commit specific offences. For instance, online marketplaces and listings services can be used by individuals to sell illegal goods or to sexually exploit adults; discussion forums and chat room services can act as spaces where suicide and self-harm is assisted or encouraged; users on online dating sites are subjected to cyberflashing; and hateful content can be shared and encouraged on online gaming services.

We note that the impact of any risk factor will depend on the combination of risk factors present on the service and how they interact. For example, a social media service offering both livestreaming and screen capture may increase the risk of CSAM on a service. Further, even when a risk factor is relevant to a number of kinds of illegal harms, the way in which it can increase risk can vary between these kinds.[25] A service's risk assessment should consider all risk factors in the round and have a good understanding of how they may affect different kinds of illegal harms, the context in which they are present, and how they interact (for further information on a service's risk assessment requirements, see the Risk Assessment Guidance and Risk Profiles).

Services with large and small user bases can increase the risks of harm to individuals

Services with large and small user bases pose risks to individuals, but often for different reasons. Large services can pose a particular risk of harm because harmful content or conduct on them can reach a large volume of people and because they sometimes attract perpetrators looking to target large volumes of users. Smaller services can pose a particular risk of harm because they may be more focused on particular interests or topics and can therefore be exploited by perpetrators

[24] Raza, S., Rahman, M., Kamawal, S., Toroghi, A., Raval, A., Navah, F. and Kazemeini, A. 2024. A Comprehensive Review of Recommender Systems: Transitioning from Theory to Practice. *ArXiv:2407.13699v1 [cs.IR].* [accessed 12 November 2024].; Salumke, T. and Nichite, U. 2022. Recommender Systems in E-commerce. *ArXiv: 2212.13910.* [accessed 12 November 2024].

[25] For example, user groups can be used by perpetrators to share CSAM or advice/tips on illegal practices OR user groups can be used by perpetrators to identify and target individuals for fraudulent activities.

looking for specific communities to target. Smaller services may also have fewer resources available to moderate content, and therefore offer more protection to a perpetrator.

For instance, a terrorist organisation may target a service with a large user base to propagate its message and increase the virality of its illegal content. The same terrorist organisation may use a smaller service to store illegal content or organise an attack due to lack of content moderation that may exist (see the Terrorism chapter for further information).

A service provider's business model and commercial profile can increase risk

In this section we provide our consideration of how a service provider's business model (revenue model and growth strategy) and commercial profile can give rise to an increased risk of illegal harms on services. In assessing the risk of harm, we are required to consider the role of the "business model" [26] of a service. Within this, we consider there are three components worth exploring:

a) Revenue model, i.e. how a service provider generates income or revenue (for instance, through advertising or subscriptions).

b) Growth strategy, i.e. how a service provider plans to expand its business (for instance, through increasing revenue and number of users).

c) Commercial profile, i.e. the size of the service provider in terms of capacity, the stage of service maturity and rate of growth in relation to users or revenue.

In and of themselves, these characteristics of a service alone do not necessarily determine the risk of harm, but they can contribute to environments on services in which a variety of illegal harms are more likely to occur. Therefore, the considerations and conclusions outlined here are relevant to every kind of illegal harm explored in the Register of Risks. Where we have further evidence that links business model characteristics very specifically with the kind of illegal harm, content or offences being described we have added it to the relevant chapter of the Register of Risks.

Revenue model

Revenue models, which ultimately dictate how online services will generate profit, can create financial incentives that – intentionally or unintentionally – lead to business decisions which prioritise revenue and profit over user safety, exposing users to an increased risk of harms.

Service providers that generate revenue in proportion to the number of service users or volume of user engagement – such as through an advertising revenue model[27] or subscription revenue model[28] – can be incentivised to design systems and features that influence user experience in a way that maximises time spent on service and engagement with content.

A system that prioritises engagement will prioritise serving users a larger volume of content or content that generates more engagement. If potentially illegal or harmful content exists on a service

[26] As part of our risk assessment duty, we must identify characteristics that are relevant to these risks of harm and assess their impact. This entails considering other aspects of a service, beyond the content presented, such as the business model. We have considered revenue models, growth strategy and commercial profile as parts of business models and the risks they pose to individuals in the United Kingdom, including how user-to-user services can be used to commit or facilitate priority offences.

[27] Service providers for which advertising is a key income stream are incentivised to report to advertisers a high user base and high user time spent, as these are key to attracting advertisers to the service. Therefore, service providers which rely on advertising revenue models have a financial incentive to promote content that drives user engagement.

[28] Subscription revenue models generate revenue in proportion to the number of paying subscribers and can create financial incentives to promote engaging content that helps attract more paying subscribers and minimise user churn.

then, unless prevented, automated systems can use this to fulfil these engagement-based goals. In this scenario, a focus on quantity of content and engagement increases the risk that a user will encounter harmful content at some point within the total volume of content they are served; that they will be served further harmful content should they engage with it once encountered; and that they will be able to find harmful content if they are searching for it. Where users themselves are also rewarded – for example, monetarily or in terms of status – as a result of the engagement their content generates, it can encourage the creation and sharing of content designed primarily with engagement as the goal. This can fuel this cycle and, where harmful or borderline content correlates with high engagement, as it has been shown to, further increases the risk of other users encountering it.

We have discussed the role of specific features and functionalities that play a role in this above, in the sub-section 'Service functionalities, features and technologies'.

As well as advertising- and subscription-based revenue models, other forms of funding are possible. Some online service providers are funded partly or wholly by donations. These may operate commercially or on a not-for-profit basis. Service providers that are funded by donations may still be incentivised to maximise user numbers and/or engagement to drive the number and size of donations, which may not necessarily incentivise services to prioritise user safety. Where a service provider is reliant on large donors, it may be influenced by those donors' priorities, which may or may not align with user safety.

In the same way in which commercial incentives may favour service design choices that increase the risk of exposure of illegal content to users, these incentives may not sufficiently support the development of systems and processes that better protect users, because investing in such measures may lead to a reduction in revenue (or profitability). One example is content moderation. It may be resource-intensive for services to accurately detect illegal or harmful content and to distinguish it from other kinds of content. Service providers may not therefore have sufficient incentive to moderate content, especially if the risk of over-blocking content (e.g. where legal content is unnecessarily taken down) could reduce user engagement or numbers, and therefore revenue.

In addition, specific features or functionalities that relate directly to how a service generates revenue can increase the risk that users encounter content that is harmful to them. One example is the ability for users to pay service providers for greater prominence of their user-generated content (e.g. boosting posts). There is a risk that bad actors could abuse this functionality by paying for harmful content (including potentially illegal content) to be boosted. Being promoted, such content reaches a wider audience, so could more easily be encountered by users and pose risks to them, if harmful.

On the other hand, the reputational risk of exposing users to illegal content could negatively affect a service's revenue in the long run, potentially giving rise to some countervailing incentives. Some users may unsubscribe from, or disengage with, services where they encounter illegal content, and business customers (for example, advertisers in advertising models[29]) or the wider industry (for example, payment providers or investors[30]) may put commercial pressure on providers to clamp

[29] For example, it was reported that recent changes to Twitter's content policies have led to a surge in harmful content on the site, and in turn, a drop in advertising revenue. Source: New York Times, 2023.

[30] Investors are one of the actors who may consider the risk of online harms and potentially influence the approach of services they may invest in. To help inform our understanding of risks, and how investors may influence the risk of online harms, we commissioned a report, 'Investors Attitudes to Online Harms – Risks, Opportunities and Emerging Trends'. This report was published alongside the Illegal Harms Consultation. Investors Attitudes to Online Harms - Risks, Opportunities, and Emerging Trends

down on such content. This may sustain some incentives to have effective measures in place to protect users from harmful content. However, to date, market incentives do not appear to have been strong enough to lead the entire online sector to put in place safety systems and processes to mitigate risk and harm on the scale required.

The link between revenue models and the risk of harm may be complex, and the extent of any trade-off between financial optimisation and user safety is likely to vary depending on the specific circumstances of each service.

Growth strategy

Growth strategies can also be associated with incentives that are in tension with user safety. Service providers may be incentivised to prioritise the use of their limited financial resources for activities and strategies aimed at growing their business (for example, marketing campaigns, research and development (R&D) activities, acquiring new assets and technologies) rather than for the development or improvement of systems and processes that protect its users from harms.

This is true especially if such systems and processes could negatively affect their growth. For instance, services whose growth strategy is aimed at increasing the user base can have a disincentive to moderate content that is harmful to UK citizens if it attracts a large number of new users quickly.

Commercial profile

The commercial profile of an organisation may influence the sophistication and relative importance of the risk management processes for a service. This, in turn, may affect the risks faced by users if it translates into a reduced capacity to protect users from encountering illegal content. For example, services that are low-capacity[31], at an early-stage[32] or have a fast-growing user base may face an increased risk of harm. For instance, all else being equal:

- Low capacity and early-stage services are less likely to have technical skills and financial resources to introduce effective risk management, compared to more mainstream services. For instance, they may have insufficient resources to adopt technically advanced automated content moderation processes (for example, automated content classifiers), or to employ a large number of paid moderators, and may rely significantly on community moderators instead. In addition, they are likely to seek growth, which may affect their incentives to have effective risk management in place.
- Service providers with a fast-growing user base may face difficulties in effectively moderating content, given the increased scale and sophistication of the moderation technologies and processes required to keep track of the user base (since the sources of risk, and kinds of harms on the service, can change quickly as the user base develops).

On the other hand, businesses with a more mature profile are likely to have larger user bases and can hence be targeted by bad actors looking to reach large populations of users with illegal content. Such services can therefore present high risks, even when they have significant resources devoted to risk management, unless appropriate systems and processes are in place to protect users from exposure to risk of harms.

As with other factors discussed in this section, the link between commercial profiles and risk of harm may be complex, and the extent of any trade-off between financial optimisation and user safety may vary depending on the specific circumstances of a service. For example, it is possible for an early-stage service to have access to substantial resources, depending on its funding context.

[31] Services with a small number of employees and/or limited revenue.
[32] A service in the initial phases of its lifecycle (for example, start-up and early growth stages).

We consider business models and commercial profiles a 'general risk factor' – relevant to all services in-scope of the Act – recognising their importance, but also acknowledging that they may affect risk of harm in an indirect way. The financial incentives and commercial context associated with a service can influence the approach to governance, service design (for example, functionalities) and systems and processes (for example, content moderation). Our Codes focus primarily on these aspects – governance, systems and processes, service design and functionalities – which can affect risk of harm more directly.

Individuals at risk of harm

Personal characteristics

Experiences of harm are incredibly personal, as is the combination and frequency of risks of illegal harm any one person encounters online. As has been noted, personal characteristics can play an important role in the extent to which people, or groups of people with shared characteristics, experience risk and harm.

One of the 'characteristics' referred to in the Online Safety Act is 'user base'. In our assessment of the causes and impacts of illegal harm online we have explored 'user base' in terms of both size and composition. As part of this, where possible we have considered various demographic and other personal characteristics of users and how this relates to the risk of harm. Where the evidence allows, we highlight personal characteristics that appear relevant to each of the kinds of illegal harm explored in the Register of Risks. Given the huge variety of personal characteristics and circumstances that have an influence on how someone experiences risk and harm online we have focused on the personal characteristics that are most often referred to in the evidence base or which are inherently related to the kind of harm or offence being explored (for example, gender differences in experience of intimate image abuse).

Vulnerability

People can be considered vulnerable to online risk and harm for a huge range of reasons. People with specific personal characteristics as described above may be considered more vulnerable than users who don't have those characteristics. But someone's personal circumstances at any one time can also play a huge role in the extent to which they are exposed to risks and an increased risk of harm as a result. Where relevant in each Register of Risks chapter, we have drawn out specific issues around potentially vulnerable users from the evidence, including considerations such as user's mental health or digital literacy.

Intersectionality

The coming together of an individual's personal characteristics may make them more vulnerable to risk and harm. For instance, a black Muslim woman may be more susceptible to harassment and hate due to the triple discrimination they face as a result of their racial, religious, and gender characteristics. Therefore, we consider intersectionality where relevant in our risk assessments to aim to understand people's experiences and risk of harm.

Media literacy's influence on risks of harm

Media literacy[33], the ability to use, understand and create media and communications in a variety of ways[34] can both contribute to and limit the risk of harm.

We broadly consider users with a strong knowledge of services and online systems, the confidence to use them adeptly, and those with a good level of critical understanding of online media to have high levels of media literacy. Those with lower levels of media literacy may struggle to navigate the online space, tend not to have good critical understanding online and find it hard to comprehend online services.

A low level of media literacy may make users more vulnerable to some forms of online harm. This could be due to a lack of awareness that means they may not recognise the harm being perpetrated until it is too late; or due to a lack of knowledge about how to raise any concerns about what is happening. Nevertheless, this does not mean or imply that the victim of harm is at fault.

Some of the illegal offences in the Act rely on a deliberate and sophisticated use of services to avoid detection. Therefore, some chapters note high levels of media literacy as a relevant risk factor for services with functionalities that may be exploited for harmful purposes. But high levels of media literacy can also be an empowering quality as it enables users to better avoid certain harms.

Ofcom's Register of Risks for illegal content

This section explains how we have conducted the analysis for our sector-wide risk assessment, the findings of which are presented in the Illegal Harms' Register of Risks. The information presented below is to help interested parties understand how we conducted our analysis, and the considerations involved in assessing the risks of harm arising from illegal content and by the use of U2U services for the commission or facilitation of priority offences.

This section covers different aspects of our approach to conducting this risk assessment. It is structured as follows:

a) **Aims and scope:** including the definition of harms and kinds of illegal harms considered
b) **Methodology:** including risk factors considered
c) **Evidence:** including considerations regarding our evidence base

Aims and scope

The Online Safety Act (the Act) requires Ofcom to carry out sector-wide risk assessments to identify and assess the risk of physical and psychological harm to individuals in the UK presented by regulated user-to-user (U2U) and search services, and to identify characteristics relevant to such risks of harm.[35]

We must publish the findings of our risk assessments in a 'Register of Risks', and then prepare 'Risk Profiles'.

The Risk Profiles are a list of characteristics of online services (which we refer to as risk factors), such as user base size, functionalities, and business model, that are likely to increase risk and indicate

[33] Ofcom is mandated to promote media literacy. Section 11 of the Communications Act 2003 [accessed 28 June 2023].
[34] Ofcom, 2023. Making Sense of Media Homepage. [accessed 14 October 2024].
[35] 'Risks of harm' refers to the harm to individuals presented by (a) content on U2U or search services that may amount to the offences listed in the Act, and (b) the use of U2U services for the commission and/or facilitation of these offences (collectively, the 'risks of harm'). 'Harm' means physical or psychological harm; we discuss physical or psychological harm as part of our assessment of the risks of harm.

which kinds of illegal harm may be more likely to occur. U2U and Search service providers are required to take account of the Risk Profiles when they carry out their illegal content risk assessment duties in the Act.

Ofcom must keep both the Register of Risks and Risk Profiles up to date. We will monitor harms and regulated services trends and will revise our Register of Risks as appropriate. In future we may expand the scope of our risk assessment if necessary. For example, as new technologies develop, and risks to online safety emerge due to the rapid innovation of the sector. This may include technologies such as immersive online virtual worlds, augmented realities, and generative artificial intelligence ('generative AI').

We will update the Register of Risks over time as new risks, risk factors, harm and evidence that underpins our understanding of these things emerges.

We will monitor future case law and associated precedents and consider updating our guidance as appropriate.

'Harm'

In the Illegal Harms Register of Risks, we consider 'harm' according to how it is defined in the Act: Harm means physical or psychological harm.[36]

As set out in the Act, harm can arise from isolated incidents, or from cumulative incidents where illegal content is repeatedly encountered by an individual, whether this is the same kind of illegal content or multiple different kinds.[37]

Harm can include circumstances of indirect harm, in which a group or individual are harmed, or the likelihood of harm is increased, as a consequence of another person encountering illegal content, which then affects their behaviours towards others.[38]

Wider societal harm

As well as exploring the harm caused to individuals, where possible in the Register of Risks we also reference the concept of wider societal harm – where there are negative collective impacts experienced by the UK public at large, associated with the specific illegal acts we cover.

We have not explored the societal harm caused by crime in its totality in the Register of Risks. But have provided information related to the scale of some kinds of illegal harm, and discussed instances of the potential societal harm where it is less inherently obvious. For example, where the prevalence and accessibility of (illegal) non-consensual intimate imagery online can normalise intimate image abuse[39]; or where society as a whole might be impacted by a foreign state seeking to manipulate how UK citizens participates in an electoral event.[40]

---

[36] Section 234(2) of the Act.

[37] See section 234(4) of the Act, which states: *"References to harm presented by content, and any other references to harm in relation to content, include references to cumulative harm arising or that may arise in the following circumstances— (a) where content, or content of a particular kind, is repeatedly encountered by an individual (including, but not limited to, where content, or a kind of content, is sent to an individual by one user or by different users or encountered as a result of algorithms used by, or functionalities of, a service); (b) where content of a particular kind is encountered by an individual in combination with content of a different kind (including, but not limited to, where a kind of content is sent to an individual by one user or by different users or encountered as a result of algorithms used by, or functionalities of, a service)."*

[38] As set out in Section 234(5) of the Act

[39] See the Intimate Image Abuse chapter

[40] See the Foreign Interference Offence chapter

Kinds of illegal harm considered

The relevant offences considered in our risk assessment are:

- Priority offences or priority illegal content, which include terrorism offences, offences related to CSEA and other priority offences. These are detailed in the Risk Assessment Guidance and Risk Profiles. So-called **'inchoate offences'**[41] are also treated as priority offences.
- **Relevant non-priority offences**[42] including the **Communications offences** (Part 10): false communications offence, threatening communications offence, offences of sending or showing flashing images electronically ('epilepsy trolling'), and offence of sending etc photograph or film of genitals ('cyberflashing'); and self-harm offences.

To make our assessment as accessible as possible, we have grouped these offences in the Register of Risks into kinds of illegal harms. This helps our analysis bring out risks that are similar in nature across the group of offences. However, within each grouping we sometimes refer to individual offences where appropriate, for example where a particular observation or evidence is relevant only to specific offences.

The kinds of illegal harm are:

1. Terrorism
2. Child Sexual Exploitation and Abuse (CSEA)
3. Grooming
4. Child Sexual Abuse Material (CSAM)
5. Hate
6. Harassment, stalking, threats and abuse
7. Controlling or coercive behaviour
8. Intimate image abuse
9. Extreme pornography
10. Sexual exploitation of adults
11. Human Trafficking
12. Unlawful Immigration
13. Fraud and financial services offences
14. Proceeds of crime
15. Drugs and psychoactive substances
16. Firearms, knives and other weapons
17. Encouraging or assisting suicide
18. Foreign interference
19. Animal cruelty
20. Non-priority offence - Epilepsy trolling
21. Non-priority offence - Cyberflashing

[41] As explained in Overview of illegal harms, 'inchoate offences' include assisting someone else to commit a priority offence, encouraging someone else to commit a priority offence, attempting to commit a priority offence or conspiring to commit a priority offence.

[42] Referred to in the Act as 'other offences', they are all offences under UK law that are not priority offences, where (a) the victim or intended victim of the offence is an individual (or individuals); (b) the offence is created as a result of the Act, another Act, an order of Council or other relevant instruments; (c) the offence does *not* concern the infringement of intellectual property rights, the safety or quality of goods, or the performance of a service by a person not qualified to perform it; and (d) the offence is *not* an offence under the Consumer Protection from Unfair Trading Regulations 2008.

22. Non-priority offence - Encouraging or assisting self-harm
23. Non-priority offence - False communications
24. Non-priority offence - Obscene content showing torture of humans and animals (the s.127(1) offence)
25. Non-priority offence - Threatening communications

'Illegal harm'

Within the Register of Risks we refer to 'illegal harm' because our analysis is focused on the risk of harm – psychological or physical – which can occur from a user encountering any illegal content, or from a U2U service being used for the commission or facilitation of an offence. However, in other parts of our guidance for service providers we refer to 'illegal content' rather than 'illegal harms'. This is because to meet the requirements of the illegal content risk assessment set out in the Act, service providers should assess the risk of harm *arising from content* that amounts to offences contained within our 17 kinds of illegal harm (1-17 above), and other non-priority offences (20-25 above).

Relationship with Protection of Children

Our Illegal Harms Register of Risks focuses on offences set out in the Online Safety Act. Content that amounts to one or more of these offences is 'illegal content', and a full breakdown of what content could amount to each offence can be found in the Illegal Content Judgements Guidance (ICJG). Both adults and children are at risk of harm from illegal content.

As well as this illegal content, the Act also defines different kinds of content that is not illegal but is harmful to children. Our Children's Register of Risks[43] is focused only on this content, which is split into three categories: Primary priority content; Priority content; and 'non-designated content' which is neither of the previous two, but which still presents a material risk of significant harm to an appreciable number of children in the United Kingdom.

There are some similarities between kinds of illegal content and content that is harmful to children. For example, content that amounts to the offence of encouraging or assisting suicide or attempted suicide is illegal. Meanwhile, content that encourages, promotes, or provides instructions for suicide is classed as content harmful to children, regardless of whether it meets the threshold for being classified as illegal. Because we cannot always definitively draw the line between illegal content and content that is harmful to children – such as within our Register of Risks chapters on encouraging or assisting suicide, encouraging or assisting serious self-harm, hate, harassment, stalking threats and abuse – we have sometimes drawn on evidence that could be describing both kinds of content as we feel it still plays an important role in demonstrating how offences manifest online and the risks of harm. Similarly, the Children's Register of Risks draws on some evidence that could be referring to illegal content.

Where relevant, we have highlighted in the Illegal Harms Register of Risks chapters where the content described by evidence may be broader than the strict definitions of illegal content.

[43] The draft Children's Register of Risks was published in our May 2024 Protection of Children Consultation.

Methodology

Understanding service characteristics as risk factors

The Act requires Ofcom to take into account how the characteristics of a service may give rise to risk. The Act defines 'characteristics' broadly as including a service's **functionalities, user base, business model, governance and other systems and processes.** We consider these characteristics both individually and, where relevant, in combination. We explain here how we have analysed characteristic risks to help individuals navigate the Register of Risks and Risk Profiles.

These characteristics form the basis of the analysis within our Register of Risks and the Risk Profiles.

Most of the characteristics referenced in the Act are not specifically defined. We recognise that given the diversity and range of services in scope of the regime, many services are likely to define some of these concepts differently. We have set out the definitions we have used to conduct our sector-wide risk assessment in the Glossary (Annex). Where possible, we have also used these terms consistently across the other regulatory documents.

The list of characteristics in the Act is not exhaustive, so it is open to Ofcom to identify other relevant characteristics. We consider our evidence justified including three additional service characteristics that can give rise to risk: **service type**, **recommender systems** and **commercial profiles.**

a) There is some evidence to suggest that certain **service types** with common features and functionalities, are more likely to be used to commit and facilitate some offences. We have therefore identified some service types as a driver of risk, although we recognise it has limitations in offering a comprehensive and robust picture of what drives risk across all services.
b) We have also identified **recommender systems** as a relevant characteristic because of the key role they play in determining what content users see and engage with, therefore contributing significantly to a user's experience of a service. Recommender systems can be used in many ways which can influence how a user might experience risk of harm on a service. Most commonly this includes content recommender systems designed for the curation of content feeds, and network recommender systems that are used to recommend other users to follow/befriend.
c) We have also included **commercial profiles** as our evidence showed that services with certain commercial profiles are likely to have weaker risk management, which can make them targets for perpetrators.

We recognise that not all characteristics are inherently harmful; we therefore use the term '**risk factor**' to describe a characteristic for which there is evidence of a risk of harm to individuals. For example, a functionality like livestreaming can be and is used by many people for benign purposes (for example, it is a very popular means by which people watch streamers playing videogames) but evidence has shown that it can be abused by perpetrators; when considering specific offences such as terrorism or CSEA, a functionality like livestreaming can give rise to risk of harm or the commission or facilitation of an offence.

Important distinctions in the Register of Risks

'Advertising' vs. 'posting goods and services for sale'

We have made a distinction between two ways in which goods or services may be promoted on a service. This is because in some cases the service provider is generating revenue from this activity: where a fee is paid to the service provider in order for goods or services to be promoted or

'advertised' on the service. In other cases, users of a service may be advertising goods or services for sale, but the service provider generates no revenue from this activity. Therefore, we distinguish between:

a) **Advertising** refers to paid-for advertising that generates direct advertising revenue for the service. This includes display advertising[44], classified advertising[45], and search advertising.[46] We cover this under 'business model'.[47]
b) **Posting goods and services for sale** refers to the ability for users to upload and share content that is dedicated to offering goods and services for sale on open channels of communication. Users may promote goods and services in this way, but it is distinct from 'classified' advertising because users do not pay for the content to be shared – although they, or buyers, may pay a fee in a different way, such as a percentage of the cost or a flat platform fee paid per sale. Therefore, it is not designed to generate direct advertising revenue for the service as classified advertising does. We cover this under 'functionality'.[48]

The reason this distinction is important is because the risks associated with how a service provider generates revenue are separate to the risks posed by the functionalities provided to users and the different ways they might be used.

Size of a service

We also use two different ways to measure service size. Although they can sometimes be correlated, it is important to distinguish them in the risk assessment because of how they might increase risks of harm to individuals.[49]

a) **Services with a large user base**: refers to services which has an average user base of 7 million or more per month in the UK.
b) **Services with a small user base**: refers to services with a small number of monthly UK users
c) **High-capacity services**: refers to services with a large number of employees and/or revenue
d) **Low-capacity services**: refers to services with a small number of employees and/or revenue

[44] 'Display advertising' is where advertisers pay to display their advertising on an online service. It can appear in a variety of formats such as banner-style adverts (e.g. a banner advert at the top of a page in the Guardian), video advertising (e.g. a video ad appearing on Mumsnet or within a YouTube video), 'native' advertising (e.g. an ad for a sponsored product appearing on a Facebook feed) and sponsored content (e.g. a sponsored article on holidays in Italy in The Sunday Times).
[45] 'Classified advertising' is where advertisers (who can be service users) pay to list specific products or services on an online service serving a market. The ad is listed under various headings and is grouped entirely in a distinct section away from display advertising. For example, an ad to sell a car in a dedicated section for car listings or advertising job opportunities under a dedicated category for job offers.
[46] 'Search advertising' is where an advertiser pays for its advert to appear within a user's search results on a search engine (e.g. on Bing, Yahoo or Google); the paid for ad will appear alongside the search engine results.
[47] This is covered under advertising revenue models. 'Boosted posts', where users pay to amplify their content, will be captured under the analysis of business model as 'transaction fees' within our consideration of revenue models.
[48] This is also sometimes considered under the umbrella of 'organic advertising'.
[49] For example, risks of harm from services with a large user base are related to higher reach, while risks from low number of employees/revenue (low-capacity service) are related to limited financial and technical ability to manage risk.

Ofcom's approach to our risk assessment

Service characteristics

In our Register of Risks we have assessed the risks of harm associated with the specific characteristics of a service.

The **characteristics** of a service as set out in the Act include any aspect of a service, including its functionalities, user base, business model, governance, and other systems and processes.[50]

- **Functionalities** is an umbrella term for the front-end features of a service that are visible to users. For U2U services, functionalities are defined as features that enable interaction between users. Functionalities for search services are defined as features that enable users to search websites or databases, as well as features that make suggestions relating to users' search requests. The Act includes a non-exhaustive list of functionalities[51], and the Ofcom risk assessment has also considered several other relevant functionalities in addition.
- **User base** refers to the users of a service. The Ofcom risk assessment has considered the size of a service's user base and user base demographics. It includes consideration of both registered and non-registered users of a service.[52]
- **Business models**, in a broad sense, refers to the ways in which a business operates to achieve its goals. For the purposes of the analysis in this Register of Risks, we adopt a narrow definition that includes revenue model and growth strategy.[53] 'Revenue model' refers to how the service generates income or revenue (for instance, through advertising or subscriptions). 'Growth strategy' refers to how the service plans to expand its business. For instance, through increasing revenue and number of users. Also see 'commercial profile' below which was not specified in the Act but which we have added.
- Governance, systems and processes (GSP):

[50] These characteristics are specified in section 98(11).

[51] Section 233: For U2U: (a) creating a user profile, including an anonymous or pseudonymous profile; (b) searching within the service for user-generated content or other users of the service; (c) forwarding content to, or sharing content with, other users of the service; (d) sharing content on other internet services; (e) sending direct messages to or speaking to other users of the service, or interacting with them in another way (for example by playing a game); (f) expressing a view on content, including, for example, by— (i) applying a 'like' or 'dislike' button or other button of that nature, (ii) applying an emoji or symbol of any kind, (iii) engaging in yes/no voting, or (iv) rating or scoring content in any way (including giving star or numerical ratings); (g) sharing current or historic location information with other users of the service, recording a user's movements, or identifying which other users of the service are nearby; (h) following or subscribing to particular kinds of content or particular users of the service; (i) creating lists, collections, archives or directories of content or users of the service; (j) tagging or labelling content present on the service; (k) uploading content relating to goods or services; (l) applying or changing settings on the service which affect the presentation of user-generated content on the service; (m) accessing other internet services through content present on the service (for example through hyperlinks).
For Search: (a) a feature that enables users to search websites or databases; (b) a feature that makes suggestions relating to users' search requests (predictive search functionality).

[52] The Act makes clear that 'it does not matter whether a person is registered to use a service' for them to be considered a 'user' (section 227 of the Online Safety Act). The Act is only concerned with the number of 'United Kingdom users' of the service, so where the user is an individual, they count as a user only where they are in the United Kingdom; similarly, where the user is an entity, they count only where they have been formed or incorporated in the United Kingdom (section 227(1) of the Online Safety Act 2023).

[53] 'Business model' can be defined more widely to describe the way in which a service creates value to its users (value proposition), how it delivers this value to users, and how it captures value for itself. However, we adopt a narrow definition in the risk assessment to avoid overlap with the other risk characteristics. This does not affect the overall risk assessment as risk factors that would have been identified under the broader definition are captured elsewhere.

- **Governance** refers to the structures that ensure the adequate oversight, accountability, and transparency of decisions within a service which affect user safety. This is in relation to organisational structure as well as product and content governance.
- **Systems and processes** refer to the actions taken by a service, including procedures to mitigate the risk of harm arising from illegal content being encountered, such as human moderators and automated systems or processes.

We also consider other characteristics that are not specified in the non-exhaustive list of characteristics in the Act, but for which there is evidence showing a relationship with the risk of harm to individuals. These include:

- **Service type.** In general, this refers to the nature of the service,[54] and includes, for example, social media services and private messaging services.
- **Recommender systems.** Refers to information retrieval systems that determine the relative ranking of suggestions made to users on a U2U service. These include systems that recommend either content (content recommender systems) or other users (network recommender systems).
- **Commercial profile.** Refers to the size of the service in terms of capacity (i.e. revenue and/or number of employees), the stage of service maturity and rate of growth in relation to users or revenue.

Within the analysis for the Register of Risks, where we find evidence of a relationship between a characteristic of a service and a harm, we consider the characteristic to be a '**risk factor**'. As such, risk factors are specific characteristics of a service which Ofcom has identified as being associated with a risk of one or more kinds of illegal harm.[55] For instance, direct messaging is a functionality that has been identified as a risk factor for some offences.

These characteristics and the associated risk factors are broad and complex in scope. To make our assessment as accessible as possible, we sometimes group risk factors that are similar in nature or increase the risks of harm in a similar way. For example, functionalities such as direct messaging and video calling have been grouped under 'user communication' because they allow users to communicate with one another in a similar way. However, they are still considered to be separate risk factors and we have assessed them accordingly.

Further information on this, including the full list of the most prominent, and potentially harmful, risk factors, is included in the Risk Profiles in the Service Risk Assessment Guidance and Risk Profiles. More information and definitions of terms used throughout this Register of Risks can be found in the Glossary (in the Annex of this document).

---

[54] Certain kinds of services or 'service types' have been selected because our evidence suggests that they can be used to facilitate or commit relevant offences.

[55] Terrorism offences, Child Sexual Exploitation and Abuse (CSEA) offences (Grooming; Child Sexual Abuse Material (CSAM)), Encouraging or assisting suicide (or attempted suicide) or serious self-harm offences, Harassment, stalking threats and abuse offences, Hate offences, Controlling or Coercive Behaviour (CCB) offence, Drugs and psychoactive substances offences, Firearms and other weapons offences, Unlawful immigration and human trafficking offences, Sexual exploitation of adults offences, Extreme pornography offence, Intimate image abuse offences, Proceeds of crime offences, Fraud and Financial services offences, Foreign Interference Offence, Communication offence - *False communications offence*, *Threatening communications offence*, *Offences of sending or showing flashing images electronically ('epilepsy trolling')*, *Offence of sending etc photograph or film of genitals ('Cyberflashing')*. For further information, refer to Chapter 5: Evidence and methodology for conducting our risk assessment.

How risk factors associated to characteristics have been identified

We used the following questions as a guide to identify the risk factors relevant to each group of characteristics:

- **Service type**: What type of service (or aspects of it) can lead to a higher risk of harms to individuals from different offences?
- **User base**: Who is using the service? How can user demographics influence which groups of users may experience or perpetrate harm and the ways in which this happens? How does the size of a user base affect risk?
- **Functionalities and recommender systems**: How can the way in which the service enables users to interact or search lead to higher risks of harm to individuals?
- **Business model and commercial profile**: How can the way in which the service achieves the goals of its business model and growth strategy lead to higher risks of harm to individuals? How can its commercial profile (capacity and maturity) affect its ability to manage risks?[56]

We acknowledge that some of the risk factors, which the evidence has demonstrated are linked to a particular kind of illegal harm, can also be beneficial to users. This can be in terms of the communication that they facilitate, or in some cases fulfilling other objectives, such as protecting user privacy.

For instance, end-to-end encryption guarantees a user's privacy and security of messages but makes it harder for service providers to moderate for illegal content. Similarly, the creation of an anonymous user profile appears to embolden user behaviour by providing users with a sense of protection, and confidence that they will not be held accountable for their actions online. This encourages some users to engage in behaviour, or post content, which they would not do if their real identity was recognisable. For example, there is evidence that hateful content targeting race or sexual orientation is more likely to be posted anonymously on some services. But at the same time, anonymous profiles allow users to question and criticise those in power without fear of repercussion, allows freedom of expression and protects a user's right to privacy.

While livestreaming can be a risk factor for several kinds of illegal harm as it can allow the real-time sharing of illegal content, it also allows for real-time updates in news, providing crucial information to a wide range of individuals.

These considerations are a key part of the analysis underpinning our Codes measures.

Evidence

Our risk assessment process has consisted in identifying and analysing a repository of quality assured evidence of over 1000 individual sources. We have considered responses from our July 2022 call for evidence, our November 2023 Illegal Harms Consultation and August 2024 Illegal Harms Further Consultation, as well as relevant Ofcom research, academic papers from a range of disciplines, government bodies including law enforcement, third-party sources and information from charities and other non-government organisations. Given the wide range of third-party evidence that we are relying on, we have taken steps to ensure that our evidence sources are reliable. In particular, we

[56] For more information on the role of advertising and service size, see the annex of this document, chapter 6W: Context to understand risk factor dynamics.

have considered the evidence by reference to the following criteria: method, robustness, ethics, independence and narrative.[57] [58]

For the purpose of our risk assessment, we have identified a list of specific characteristics of services that we considered may be relevant to the risks of different kinds of illegal harms. We have then assessed any relevant evidence of whether and how particular kinds of illegal harm are impacted by the presence or absence of those characteristics, either individually or in combination.

We have also engaged with external stakeholders including law enforcement and specialist agencies dealing with online threats to ensure we represent harms accurately.

Most of the evidence reflects the experiences of UK users. However, for some offences, we have used research from other parts of the world where we felt it helps understand online experiences, either by complementing any UK evidence available, or providing additional insights in cases where there was no UK evidence.

Evidence base

Despite the extensive review of evidence, there remain gaps for some offences; in particular, there are gaps in the evidence that link the characteristics the Act requires us to assess against the kinds of illegal harms.[59]

At present, we hold less evidence about risk on search services compared to U2U services – there is less publicly available information about how they operate, and about the presence of illegal content in search results that can cause harm to individuals on these services.

The amount of available evidence for specific kinds of illegal harm and offences is also varied. We have found it to be limited for some kinds of illegal harm (for example, extreme pornographic content offence), and for how services were used for the commission or facilitation of certain priority offences (for example, unlawful immigration, human trafficking, and firearms, knives and other weapons offences). We do not take this as an indication that the content or offences do not cause harm online, but as a reflection of the lack of reliable evidence at this time. We are also aware of the ethical and legal limitations to conducting research into certain kinds of illegal harm; research in those cases often focused on qualitative information instead. In some cases, we have been able to support our understanding of these harms by engaging with law enforcement and other specialist agencies.

Where evidence is limited, we have used our judgment and expertise about specific harms to draw conclusions where we think this can help services to identify potential risks. We have also relied on some evidence that is about content or conduct that is broader than the specific offences, where we consider that this is nevertheless likely to be relevant to those offences. We have signposted this in the relevant parts of the Register of Risks. Due to the fast pace of technological changes and the

---

[57] 'Method' examined the strengths and weaknesses of the methodology for that particular topic, such as whether appropriate data collection methods were used. 'Robustness' considered both the size and coverage of the sample, and quality of analysis – for example, how missing data values were accounted for. 'Ethics' refers to how well ethical considerations were addressed in the study, such as how personal data was handled. 'Independence' examined the origins of the research and whether any stakeholder interests might have influenced findings. 'Narrative' refers to the commentary within the report and whether conclusions are sufficiently backed by the research, and whether there is a clear distinction between the findings and the interpretation.

[58] Some of the evidence used in this Register was published in a response to the development of the Act and other relevant legislation. These sources may have had aims or ambitions associated with the development of legislation. Moreover, some of the evidence used in this risk assessment comes from experts in their field, who may have developed their expertise while in the former employment of online services.

[59] For example, our evidence base assessing governance, systems and processes and illegal harms is under-researched in some areas. We have therefore used different types of research and supporting evidence in this analysis.

speed at which risks of harm can manifest online, some of the evidence used within our risk assessment has come from non-traditional research sources; this timely evidence may not have the traditional levels of methodological and sampling rigour and peer reviewing that more traditional research sources have. This includes the use of videos and podcasts, as well as the use of investigative journalism.

Lastly, we would highlight two important observations regarding the evidence of 'types' of service. First, some of the research-based evidence we refer to relates to specific services. We have included this evidence because it provides insights about particular risks that we consider have more general application. Its inclusion should not be seen as a judgement about the online safety practices of those specific service providers. Second, we do not have specific evidence relating to all types of U2U services. There is more research available - including on risks of harm to individuals - about large social media sites, gaming sites, and services that publish public information that can be analysed. Where appropriate, we have made reasonable inferences about the risks that may arise on other services where we do not have specific evidence about that service type.

Using the Register of Risks

The detailed research and analysis set out in our Register of Risks is intended to help services comply with their obligations under the Act. Over the subsequent 25 chapters, we set out our full risk assessment for illegal content on U2U and search services and consider the use of regulated U2U services to commit or facilitate priority offences.

The Register of Risks is split into three parts:

a) The first part **user-to-user services** (chapters 1 to 23) identifies characteristics of U2U services that may lead to increased risks of harm to individuals in relation to each of the kinds of illegal harm covered by the Act. This includes functionalities and recommender systems, user base, business models and commercial profiles. We consider both the risk of harm presented by the dissemination of illegal content on a U2U service, as well as the use of these services for the commission and/or facilitation of each kind of illegal harm. The risks from each kind of illegal harm on U2U services are explored in their own chapters.
b) The second part **search services** (chapter 24) identifies the characteristics which can increase the risk of harm to individuals on search services. For search services, we only consider illegal content and not the use of a search service for the commission or facilitation of an offence (as per the Act's requirements). We consider all of the kinds of illegal harm together.
c) The third and final part (chapter 25) explores how the **governance, systems and processes** of a U2U or search service may lead to an increased risk of harm. We have identified two general scenarios where risk can arise from these areas themselves: (a) inadequate governance and/or other systems and processes currently in place within regulated services; and/or subsequently (b) an absence of such governance and other systems and processes.

We also have an Annex which includes a glossary of terms used throughout the Register of Risks.

User-to-user services

This part of the Register of Risks presents a detailed analysis of the kinds of illegal harm, and their associated risks, on user-to-user (U2U) services.

This part includes 23 chapters. 18 cover the 'kinds of illegal harms' into which we have grouped all the priority offences in the Act, and a further 5, each covering a relevant non-priority offence.

The chapters on kinds of illegal harm distinguish each priority offence where the evidence permits, and consider risks of harm to users more widely, where justified. For example, the Terrorism chapter includes a number of priority offences relating to terrorism. Within it, we have evidence pointing to the risks of harm relating to several of the terrorism offences listed, while other evidence may point to the particular harm of one specific terrorism offence. We have analysed evidence for the other relevant non-priority offences in the same way, using the same structure as the other chapters.

In each chapter we have considered evidence from a variety of sources, including information provided by services, academic literature, third-party research, civil society in general and Ofcom's own research.

Each chapter is structured as follows:

a) Summary of the chapter, including important risk factors identified.

b) Introduction to the harm and the relevant offences covered.

c) How the offences manifest online. This reviews the presence of the harm online and the risks of harm that users may experience. This will help a service understand the context for the harms and the particular risks a service should be aware of.

d) Evidence of risk factors. The evidence to form the basis of our analysis is presented for each characteristic: service type, user base, functionalities and recommender systems, and business models and commercial profiles. This final section will allow services to develop a better understanding of how specific characteristics relate to, and impact, the risks of harm.

U2U service types

We refer to U2U service types that we expect to be recognisable to both users and businesses, to illustrate how harms can manifest online and how the characteristics of a service can affect the risks of harm to individuals.

The U2U service types below should not be taken to be a definitive view of the services (or parts of services) that may be in scope of the Act or a classification that sets expectations about a service provider's risk assessment. It is for services to assess themselves and seek their own independent advice to enable them to understand and comply with the Act. For more, please refer to the 'Overview of regulated services':

- **Social media services:** Social media services connect users and enable them to build communities around common interests or connections.
- **Video-sharing services:** Video-sharing services allow users to upload and share videos with the public.
- **User-to-user pornography services:** Service type whose principal purpose is to disseminate user-generated pornography.
- **Discussion forums and chat room services:** Discussion forums and chat rooms generally allow users to send or post messages that can be read by the public or by an open group of people.
- **Marketplaces and listings services:** Marketplaces and listings services allow users to buy and sell their goods or services.

- **Dating services:** Dating services enable users to find and communicate with romantic or sexual partners.
- **Gaming services:** Gaming services allow users to interact within partially or fully simulated virtual environments.
- **Messaging services:** Messaging services are typically centred around the sending and receiving of messages that can only be viewed or read by a specific recipient or group of people.
- **File-storage and file-sharing services:** File-storage and file-sharing services are services whose primary functionalities involve enabling users to store digital content and share access to that content through links.
- **Information-sharing services:** Information sharing services are primarily focused on providing user-generated informational resources to other users.
- **Fundraising services:** Fundraising services typically enable users to create fundraising campaigns and collect donations from users.
- **Payment services**: Financial payment providers often have websites or applications that enable users to send and receive money.[60]

Recent developments such as GenAI can also be relevant when considering service types.

We will continue to monitor the U2U landscape with the expectation that new types of services and research showing a risk of harm associated with them will emerge.

[60] These services can sometimes allow users to share user-generated content such as messages.

1. Terrorism

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for terrorism offences: how harm manifests online, and risk factors

Terrorism is considered a violent action or threat of action, designed to influence a government or intimidate the public and advance a cause. Online terrorism content is any content made available to others online, which can encourage or promote terrorism. Although online terrorism content is not widespread on user-to-user (U2U) services, the impact on individuals and communities can be substantial, both physically and mentally.

Service type risk factors:

Terrorist content encountered by UK users does not rely on a single service but on many services and their associated functionalities. We have found that there are often cross platform sharing of terrorism content; from being first posted on smaller U2U services and then linked to larger, higher-reach services and vice versa.

A wide range of types of U2U services are known to be used by terrorist actors. **Social media services** are particularly relevant to the perpetration of this harm because of their reach and popularity. Terrorist content is also often identified on **file-storage and file-sharing services.** Similarly, **file-storage and file-sharing services** have been identified specifically as a risk factor in facilitating the creation of 3D-printed firearms.

Gaming services have also been used by terrorists as recruitment and training tools, while **marketplaces and listing services** can be used to raise and collect funds.

Other services are also used to organise, recruit, fundraise and disseminate terrorism content. These include **video-sharing services, discussion forums and chat rooms, messaging services, fundraising services, and payment services**.

Our evidence suggests that services which facilitate the creation of online communities of like-minded individuals, such as in discussion forums or chat rooms, may increase the risks of harm related to terrorism. They can enable potential perpetrators and organised communities to encourage each other to share terrorism content, which may lead to an increase in the risks of harm from terrorism.

User base risk factors:

User base size can increase the risks of harm from terrorism offences. U2U services with a large user base and high reach are a risk factor because services with a large user base can enable the dissemination of terrorism content to many users, often quickly or virally. However, services with a small user base can also be used by

perpetrators to undertake more sensitive activities, such as recruitment, planning and fundraising.

Research also indicates that men are more likely to encounter radicalisation and terrorist content than women. And children are more vulnerable to becoming radicalised.

Functionalities and recommender systems risk factors:

Many of the functionalities listed are common across U2U services; in principle, a wide range of U2U services could be used for disseminating terrorism content, and this makes it harder to identify which services are especially risky, based solely on a general analysis of functionalities.

Perpetrators often use functionalities such as **posting content, commenting on content** and **hyperlinking** to share and direct users to content such as memes, and content which provides instructions related to terrorist activities. Our evidence points to these functionalities increasing the likelihood of terrorism content being shared. This, in turn, increases the risk of harm to individuals exposed to this content and could lead to incitement to commit terrorist acts, the dissemination of material such as weapons training, and the recruitment of people. **User connections** and **user tagging** also allow terrorism content to be disseminated through users' networks, especially when official pages or channels are removed.

The ability to **livestream** is a risk factor that has been used to broadcast terrorist attacks and to target groups with protected characteristics. In the past, this functionality has been abused or exploited on many occasions to incite and encourage terrorism, particularly by far-right terrorists who often seek to emulate the tactics of previous terrorists.

Any service offering **group messaging** can allow terrorists to share content in a low-friction way with like-minded people. Encrypted messaging is particularly attractive to terrorist groups as this can reduce the chance of detection.

Several other functionalities are relevant to terrorism offences. **Screen capturing or recording** increases the risks of harm by enabling users to store and disseminate extremist content. **User-generated content searching** allows individuals to easily seek out terrorist content while **content recommender systems** can increase the risk of exposure to it.

Our evidence finds that **direct, encrypted, and ephemeral messaging** are also used by terrorist actors for organisation and security purposes. **Anonymous user profiles** can also heighten the risks of harm, with users less fearful of sharing such content when they are anonymous. Users have also been shown to create **fake user profiles** by altering their usernames to avoid having their accounts blocked.

Business model risk factors:

The **capacity and maturity** of a service can contribute to risk and be exploited by perpetrators, with varying effects on users. Low capacity and early-stage services may be more vulnerably to risks of hosting terrorist content because of their

limited knowledge, resources or technical capability to moderate such content. This reflects the opportunistic nature of perpetrators to use online services to heighten the risks of harm from terrorism, for varying purposes.

Introduction

1.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to the terrorism offences listed under 'Relevant offences'; and
- The use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

1.2 We set out the characteristics of U2U services that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm; we discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the effect of cumulative harm where content of a particular kind is repeatedly encountered by an individual or encountered in combination with content of a different kind.

Relevant offences

1.3 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding terrorism offences, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 5 of the Act.

1.4 The priority offences for terrorism are the following:

- Membership in a proscribed organisation[61]
- Inviting support for a proscribed organisation
- Expressing an opinion or belief supportive of a proscribed organisation
- Arranging a meeting supportive of a proscribed organisation
- Publishing an image of the uniform of the proscribed organisation
- Terrorist fund-raising
- Use of money or property for terrorist purposes
- Possession of money or property for terrorist purposes
- Involvement in terrorist funding arrangements
- Laundering of terrorist property
- Providing weapons training
- Inviting another to receive weapons training
- Directing a terrorist organisation
- Collection of information likely to be of use to a terrorist

[61] The offences listed at points (a) to (p) refer to the offences under the following provisions of the Terrorism Act 2000: sections 11; 12(1), 12(1A); 12(2), 13(1A), 15; 16(1), 16(2); 17; 18; 54(1); 54(3); 56; 58; 58A; sections 59 to 61f.

- Publishing information about members of the armed forces etc
- Inciting terrorism outside the United Kingdom
- Use of noxious substances or things[62]
- Encouragement of terrorism[63]
- Dissemination of terrorist publications
- Preparation of terrorist acts
- Training for terrorism
- Terrorist threats relating to radioactive devices

1.5 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of some of these offences (and, in relation to offences in Scotland, being involved art and part in the commission of those offences).

1.6 For more details on the offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

1.7 If the action, or threat of action, involves the use of firearms or explosives, it will be considered 'terrorism', whether or not the action is designed to influence the government or an intergovernmental organisation, or to intimidate the public or a section of the public. 'Action' includes action outside the UK. Under British terrorism legislation, an offence is not limited to the committing of a terrorist attack, but extends to the planning, assisting and even collating information on how to commit terrorist acts.[64]

1.8 The UK Government publishes a list of proscribed terrorist organisations.[65] To proscribe an organisation, the Home Secretary must have a reasonable belief that it is currently concerned in terrorism and that it is proportionate to proscribe. The Home Secretary will make this decision having considered all relevant factors, including the specific threat a group poses to the UK. Proscription decisions require approval from both Houses of Parliament. In addition to the relevant terrorism offences in the Terrorism Acts 2000 and 2006, there are terrorism offences associated with proscription which include, but are not limited to, offences such as belonging to, or professing to belong to, a proscribed organisation in the UK or overseas.

1.9 Examples of terrorism content online include, but are not limited to, posts, text, images and videos that incite terrorist activity, instructions on how to commit a terrorist attack, connecting to people for recruitment, filming and livestreaming a terrorist attack with the intention of glorifying the act or building support for the perpetrator or proscribed organisation they associate with[66], inviting support, and the promotion of terrorism content through discussion forums and chat rooms, as well as on social media services.

---

[62] The offence listed at point (q) refers to the offence under section 113 of the Anti-terrorism, Crime and Security Act 2001.
[63] The offences listed at point (r) to (v) cover the offences under the following provisions of the Terrorism Act 2006: sections 1; 2; 5; 6; 11.
[64] The Crown Prosecution Service, 2022. Terrorism. [accessed 29 June 2023].
[65] The Home Office, 2021. Proscribed terrorist groups or organisations. [accessed 29 June 2023].
[66] Note that in the event of a livestreamed attack, service providers are unlikely to be in a position to carry out an assessment of the motives of the attacker in sufficient time, so it is much more likely the content would be 'illegal content' under a public order offence. See the ICJG chapter 'Terrorism' for more information.

How terrorism offences manifest online

1.10 This section is an overview which looks at how terrorism manifests online**,** and how individuals may be at risk of harm. To put the risks of harm to individuals from terrorism content into context, Ofcom's Online Experiences Tracker (OET) found that 8% of UK internet users aged 13 and above claimed they had experienced content that encouraged extremism, radicalisation or terrorism in the past four weeks; concern for encountering this type of content was high, with 80% (4 in 5) expressing high concern.[67] Over a 'lifetime', the chances of encountering potentially illegal content are higher, with suggestions that as many as nearly 1 in 4 adults (23%) could have, at some point, actively sought out information that *"incited or facilitated terrorism."*[68]

1.11 Numerous sources demonstrate the role U2U services have played in the sharing of and access to terrorist content, leading to a risk of harm to individuals. Our evidence shows that terrorist groups and actors use online services in a variety of ways.

1.12 Social media services are a tool for dispersing terrorism content and are a medium that terrorist organisations use to recruit, but there is analysis indicating that these are just one online arena that terrorist actors use.[69] Evidence points to terrorism content moving from larger, mainstream social media services to smaller services. A Tech Against Terrorism[70] report also talks about an ongoing migration of terrorism content to emerging video-sharing services, as *"medium-sized services increase their capability to identify and remove terrorist or violent extremist content".*[71]

1.13 Some terrorism-operated websites may not be U2U services and may therefore be out of scope of this chapter, but still form an important part of the wider ecosystem of terrorism content and can be linked to via URLs hosted on regulated services.[72] Tech Against Terrorism reported on this resurgence of terrorist-operated websites during 2021, explaining that terrorist organisations create these to further their goals.[73]

1.14 Cross-platform exploitation was also highlighted following the Buffalo terrorist attack of May 2022: "URLs pointing to smaller services or third-party sites, hosting live-streamed footage of the attack, were shared across a variety of platforms".[74] This tactic used by

---

[67] Ofcom, 2024. Online Experiences Tracker - Wave 6. [accessed 11 November 2024].

[68] In a survey with UK adults, 23% were identified as displaying information use behaviour that reflected a moderate or high likelihood of actively seeking information that incited/facilitated terrorism, and that 31% of participants exhibited information use behaviours that suggested a moderate of high likelihood of being incidentally exposed to such information. Note that this is a much broader definition of potentially terrorist content and exposure risk than the Online Experiences Tracker. Source: Schuman, S., Clemmow, C., Rottweiler, B., Gill, P. 2024. Distinct patterns of incidental exposure to and active selection of radicalizing information indicate varying levels of support for violent extremism. [accessed 30 August 2024].

[69] For example, a study conducted by the Tony Blair Institute states that *'to access extremist content via social media, you need to know where to look and, in most cases, individuals will already have been exposed to terrorist thinking through social circles offline, or they will be aware of accounts disseminating content and will actively follow them on other platforms.'* Source: Tony Blair Institute, 2016. A War of Keywords: How extremists are exploiting the internet and what to do about it. [accessed 19 September 2023].

[70] Tech Against Terrorism is an initiative launched and supported by the United Nations Counter Terrorism Executive Directorate (UN CTED) working with the global tech industry to tackle terrorist use of the internet whilst respecting human rights.

[71] Tech Against Terrorism, 2021. Trends in Terrorist and Violent Extremist Use of the Internet. [accessed 29 June 2023].

[72] Under the Act, Ofcom is required to identify and assess risks connected with regulated U2U and search services. If a standalone website does not offer either of these services, it is likely to be out of scope of regulation.

[73] Tech Against Terrorism, 2021.

[74] Ofcom, 2022. The Buffalo attack: Implications for online safety. [accessed 29 June 2023].

potential perpetrators makes it harder for a service to track down terrorism content, which can lead to it reaching a higher number of individuals.

1.15 The ability to share files for 3D-printing of prohibited weapons or weapon components is an emerging way in which U2U services can pose a risk of harm in the manifestation of terrorism offences.[75] This evolving threat from 3D-printed firearms[76] has resulted in a recent consultation launched by the Government.[77] The National Crime Agency (NCA), along with other law enforcement authorities, has highlighted the extent of the threat through recent arrests and convictions.[78] Clear links have also been seen that connects 3D printed firearms and associated online activity with extreme ideologies, particularly UK Extreme Right-Wing ideology.[79]

1.16 A similar example is the use of digital templates to print after-market components to convert semi-automatic firearms into automatic firearms.[80] It is important to identify that while there were no confirmed criminal discharges using 3D-printed firearms in 2023, it is highly likely that criminals have a growing interest in hybrid 3D-printed firearms.[81] See the Firearms, knives and other weapons chapter for information on weapons offences covered by the Online Safety Act.

1.17 Finally, there are also risks associated with the spread of generative AI. Limited evidence exists on the link between generative AI and other emerging technologies and terrorism recruitment. However, there is evidence of terrorist groups using AI to create terrorist propaganda.[82] Because generative AI lowers barriers to creating propaganda at scale, it is conceivable that the integration of generative AI tools into services could in some circumstances increase the risks of these services being used for radicalisation. We are working on building our evidence base around AI and emerging technologies to better assess the risks they pose.

---

[75] The provision of blueprints or instructions which allow another person to 3D print ('make') a firearm would be considered illegal content in relation to offences relating to the provision of instructions or training in the making or use of firearms. Refer to the ICJG (Terrorism offences chapter) for further details.

[76] The 2024 National Strategic Assessment from the National Crime Agency (NCA) stated "*there were no confirmed criminal discharges using 3D-printed firearms in 2023, although it is highly likely that criminals have a growing interest in hybrid 3D-printed firearms.*" Source: NCA, 2024. Criminals still want to acquire and use original lethal purpose weapons but they are finding them more difficult to obtain. [accessed 22 October 2024]; Global Network on Extremism and Technology (GNET) (Basra, R.), 2022. The Future is Now: The Use of 3D-Printed Guns by Extremists and Terrorists. [accessed 22 October 2024].

[77] Potential new laws to criminalise the making, supply and possession of items strongly suspected to facilitate serious crime – such as digital templates for 3D-printing firearms components. Source: Home Office, 2023. Consultation document (accessible). [accessed 20 September 2023].

[78] For example, in 2023 and 2024, several men have been convicted for possession of semi-automatic firearms, printed from files available online. For example: Laversuch, C., 2023. Pair jailed for 3D-printed sub-machine gun plot, BBC News, 16 May. [accessed 22 October 2024]; Staffordshire Police, 2024. Man who made guns using 3D printer jailed for more than 10 years. [accessed 22 October 2024].

[79] For example, several British men have been convicted of terrorism offences in recent years in cases where 3D printed firearms were linked to the offenders engagement with extreme right-wing ideology. Sources: Counter Terrorism Policing, 2024. Portsmouth man sentenced for nine years and six months for terrorism offences. [accessed 18 November 2024]; The Independent (Dearden, L.), 2022. 'Fascist cell' convicted of terror and firearm offences after trying to make 3D-printed gun, 29 March. [accessed 18 November 2024].

[80] National Crime Agency response to the Illegal Harm November 2023 Consultation.

[81] NCA, 2024.

[82] Tech Against Terrorism, 2023. Early Terrorist Adoption of Generative AI [accessed 12 June 2024].

Risks of harm to individuals presented by online terrorism offences

1.18 Terrorism content on U2U services creates risks of harm to individuals in that it can recruit a user to the terrorist's cause, which increases the risk of committing offences online (and offline). Research conducted by the Prison and Probation Service on individuals who had been convicted of an 'extremist offence'[83] shows that the internet[84] plays an increasingly prominent role in radicalisation processes, in line with wider society's increased use of the internet.[85]

1.19 Terrorism content on U2U services can also spread terrorist messages, which increases the risk that individuals are exposed to content that they may find harmful or may be harmful to them.

1.20 The objectives of terrorist propaganda may include the use of psychological manipulation to undermine an individual's belief in certain collective social values, or to propagate a sense of heightened anxiety, fear or panic in a population or subset of the population. "*This may be achieved through the dissemination of disinformation, rumours, threats of violence or images relating to provocative acts of violence. The intended audience may include direct viewers of content, as well as those affected by the potential publicity generated by such material*".[86]

1.21 The impact of encountering content that encourages or promotes terrorism can be immense and felt not only by individuals but also across families and communities. Evidence shows that explicit threats of violence and the depiction of violence, including concerning the use of weapons, which can be disseminated online, can induce anxiety, fear or panic in a population or subset of the population.[87] For further evidence on the risks of harms to individuals that can occur from threats of violence, see the Harassment, stalking, threats and abuse chapter.

---

[83] Defined as "*any offence committed in association with a group, cause, and/or ideology that propagates extremist views or actions and justifies the use of violence or illegal conduct in pursuit of its objective"* Source: HM Prison and Probation Service, 2019. Exploring the role of the Internet in radicalisation and offending of convicted extremists. [accessed 3 July 2023].

[84] The research states that information was coded relating to internet activities and behaviours commonly associated with online radicalisation. The following variables were coded: 1. Learnt from online sources. 2. Interact with co-ideologues online. 3. Generate their own extremist propaganda online. 4. Provision of material support online. 5. Access to specific extremist websites. 6 Use of open social media platforms. 7. Use of email/standard chat applications. 8. Use of encrypted applications. It is therefore possible that activity conducted on out-of-scope services has been captured as part of this research. Source: HM Prison and Probation Service (Kenyon. J, Binder. J, and Baker-Beall, C.), 2021. Exploring the role of the Internet in radicalisation and offending of convicted extremists. [accessed 3 July 2023].

[85] It is worth noting that the scope of the research extended beyond regulated services and included websites that are unlikely to offer any in-scope user-to-user or search services, and email service providers. However, many services included in the report are likely to be within scope of regulation and hence the insights highly relevant for an assessment of the risks of harm from terrorism offences. Furthermore, other evidence in this chapter also points to websites with terrorism content being part of the wider online ecosystem and at times being found or accessed through the sharing of links on in-scope service. Radicalisation was found to take place primarily online, particularly between 2019 and 2021. However, it is currently unclear to what extent the Covid-19 pandemic and associated restrictions accounted for this. The primary method of radicalisation for individuals who were convicted between 2015 and 2017 was as follows: internet: 17 individuals (27%); face-to-face: 11 individuals (17%) and hybrid: 36 individuals (56%). The report states that despite evidence suggesting the increasing prominence of the internet in radicalisation processes, it cannot be concluded that the online domain is simply replacing the offline domain, as offline influences such as previous involvement or conviction for non-terrorism offences featured at least to some extent for most convicted extremists in the dataset. Source: HM Prison and Probation Service report, 2021. [accessed 19 September 2023].

[86] United Nations Office of Drugs and Crime, 2012. The Use of the Internet for Terrorist Purposes. [accessed 29 June 2023].

[87] United Nations Office of Drugs and Crime, 2012.

1.22 Some harms relating to terrorism have a nation-specific dimension, including those linking to sectarianism and paramilitary activity in Northern Ireland. Research from the Institute for Strategic Dialogue (ISD) demonstrates that social media services are regularly used to amplify sectarian messaging or for proscribed paramilitary groups to share propaganda, recruit and mobilise.[88]

Evidence of risk factors on user-to-user services

1.23 We consider that the risk factors below are likely to increase the risks of harm relating to terrorism offences. This is also summarised at the start of this chapter.

Risk factors: Service types

1.24 Research indicates that a range of U2U services can be used to commit or facilitate offences related to terrorism. Specifically, the evidence we have reviewed suggests that the following service types can be used to commit or facilitate these offences: social media services, video-sharing services, file-sharing services, discussion forums and chat rooms, private messaging services, online gaming services, online marketplaces and listings services, fundraising services, and payment services.[89]

Social media services and messaging services

1.25 Our evidence highlights that many terrorist organisations or terrorist actors will use larger, mainstream social media services due to their substantial reach.[90] [91] Tech Against Terrorism has reported that "*terrorist networks are increasingly attempting to operate on mainstream social media platforms by masquerading as legitimate news organisations*".[92]

1.26 Restricted spaces on social media services, as well as online closed groups in messaging services, and discussion forums and chat rooms, all help to separate in-group participants from outsiders.[93] Group chats in private messaging services with encrypted messaging are used by terrorists and violent extremists to signpost content, with lesser risk of detection and sanction by content moderation teams.[94]

[88] The Institute for Strategic Dialogue (ISD) is an organisation dedicated to reversing rising extremism worldwide. Source: Institute for Strategic Dialogue, Northern Ireland Related Terrorism (Manzi, Z.), 2024. [accessed 22 October 2024].
[89] Further information on the specific ways in which these services may be used can also be found under Risk factors: Functionalities and recommender systems.
[90] Observer Research Foundation (Saltman, E.), 2022. Identifying and Removing Terrorist Content Online: Cross-Platform Solutions. [accessed 4 July 2023].
[91] Internal 'social media strategy' documents from the proscribed terrorist organisation ISIS provide a simple example of how important the group felt it was to maintain a social media presence on large mainstream services, as far back as 2014 and earlier. Source: Berger, J.M., & Morgan, J. 2015. The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter. p.55. [accessed 12 September 2024].
[92] Tech Against Terrorism, 2021. Terrorist use of E2EE: State of play, misconceptions and mitigation strategies. [accessed 3 July 2023].
[93] Texas National Security Review (Fishman, B.), 2019. Crossroads: Counter-terrorism and the Internet. [accessed 3 July 2023].
[94] Tech Against Terrorism, 2021.

Discussion forums and chat room services

1.27 Research by the Prison and Probation Service indicates that there has been increased use of discussion forums and chat rooms, as well as social media services, for the commission or facilitation of terrorist offences.[95]

File-storage and file-sharing services

1.28 Tech Against Terrorism's transparency report is a database of verified terrorism content, *"collected in real time from verified terrorist channels on messaging platforms and apps".* The report shows that terrorism content was detected on 13 different types of services; the *"three most exploited technology types in descending order were platforms providing file sharing, archiving, and link shortening services*".[96]

1.29 Recent evidence suggests that the majority of terrorism content is identified on file-storage and file-sharing services. The use of these file-sharing services is also often combined with the use of larger 'beacon' services. 'Beacon' services act as centrally located 'lighthouses' which signpost users to where content may be found. This is often done by hyperlinking to 'content stores' such as file-sharing services. Terrorists often use these beacon services and have official channels on them which aggregate their central communications.[97]

1.30 The distribution of files containing blueprints for 3D-printed firearms online has made it easier for users to illegally manufacture firearms and components.[98] The blueprint for the first fully functional 3D-printed firearm began circulation online in 2013 and continues to be distributed widely across online communities through file-sharing services.[99] It is recognised that digital blueprints for 3D-printed firearms have been distributed via online repositories, often containing files that have garnered thousands of views and hundreds of unique downloads.[100] [101] Manufacturers use file-storage and file-sharing services to share draft proposals and then download these designs for use with a 3D printer, potentially leading to the illegal possession and supply of firearms to carry out offences related to terrorism.[102] Law enforcement has similarly identified the use of file-sharing services as a risk factor in the production of 3D-printed firearms. They are also concerned that the ease of access to these files online could support an increase in the interest in 3D-printed firearms amongst minors.[103]

---

[95] HM Prison and Probation Service (Kenyon, J., Binder, J. and Baker-Beall, C.), 2022. The Internet and radicalisation pathways: technological advances, relevance of mental health and role of attackers. [accessed 3 July 2023].
[96] Tech Against Terrorism, 2022. Terrorist content analytics platform: year one: 1 December 2020 – 30 November 2023. [accessed 3 July 2023].
[97] Tech Against Terrorism, 2023. Patterns of Online Terrorist Exploitation. [accessed 3 July 2023].
[98] In 2022 Police raided a gun factory and found 3D printed components understood to play a role in the firearms. Source: Davis, M., 2022. Haul of 3D-printed gun parts and bullets one of largest in UK, The Independent, 12 October. [accessed 24 September 2024]; Dass, R. and Mok, B., 2023. Assessing the Impact of 3D-Printed Weapons on the Violent Extremist Milieu. [accessed 24 September 2024].
[99] Global Network on Extremism and Technology (Lewis, J.), 2021. 3D-Printed Guns, Untraceable Firearms, and Domestic Violent Extremist Actors. [accessed 17 October 2024].
[100] One type of 3D-printed firearm was identified as being used in at least 18 open source reports of conflict between 2020 and 2023. Sources: Basra, R., 2023. Behind the Mask: Uncovering the Extremist Messages of a 3D-Printed Gun Designer. [accessed 11 September 2024]; Dass, R., 2023. 3D-Printed Weapons and the Far-Right: The Finnish Accelerationist Cell. [accessed 11 September 2024].
[101] Miotto, N., 2021. The Role of Online Communities in Supporting 3D-Printed Firearms. [accessed 17 October 2024].
[102] Advancements in technology have led to 3D printers becoming much more advanced. A CAD or STL file can now be sent directly to the printer via a LAN connection or Bluetooth.
[103] Counter-terror Policing response to the Illegal Harm November 2023 Consultation.

Gaming services

1.31 Gaming services can be used by terrorist organisations to recruit minors. A United Nations paper found that terrorist organisations have designed or modified online video games, intending them to be used as recruitment and training tools. [104] **Online services which allow modifications to take place, also known as 'modding', can be a risk factor for gaming services due to this phenomenon.** Such games may promote the use of violence against a state or a prominent political figure, rewarding virtual successes, and may be offered in multiple languages to appeal to a broad audience.[105]

Marketplaces and listings services, fundraising services, and payment services

1.32 Marketplaces, fundraising services and payment services are used by terrorists to raise and collect funds. These kinds of services typically allow users to directly solicit funds, sell products through e-commerce, host charitable organisations and support online payments. Our evidence shows that these are some of the primary categories of ways in which terrorist actors and terrorist organisations raise and collect funds and resources.[106] For more information on this, see the Transactions and offers section.

1.33 Law enforcement has also identified that the documentation, advice and blueprints for manufacturing 3D-printed firearms are often shared using online chat rooms or forums.[107] Enthusiasts and extremists are often made aware of file-sharing repositories through online communities used to share knowledge on effective manufacturing methodology, as well as wider beliefs or political views.[108]

Risk factors: User base

User base size

1.34 Evidence suggests that the broad reach of online services may increase the risks of harm from terrorism as it can provide terrorist organisations and supporters with a global pool of potential recruits, facilitating the recruitment process.[109] This same broad reach can also increase the risks of harm experienced by users due to the potential wide dissemination of harmful content. Services with a large user base can therefore be used by terrorist actors.

1.35 Services with a small user base and less reach can also be used by terrorist actors but for different reasons. For example, while services with a large user base may be used to attract and draw individuals into the group through influence tactics and dissemination of propaganda, smaller services can be used by perpetrators to undertake more sensitive activities, such as recruitment, planning and fundraising.

User base demographics

1.36 The following section outlines the primary evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

---

[104] United Nations Office of Drugs and Crime, 2012. The Use of the Internet for Terrorist Purposes. [accessed 29 June 2023].
[105] United Nations Office on Drugs & Crime report, 2012.
[106] United Nations Office on Drugs & Crime report, 2012.
[107] National Crime Agency, 2023.
[108] Dass, R., 2023. 3D-Printed Weapons and the Far-Right: The Finnish Accelerationist Cell. [accessed 11 September 2024].
[109] United Nations Office of Drugs and Crime, 2012. The Use of the Internet for Terrorist Purposes. [accessed 29 June 2023].

1.37 Data suggests that user base characteristics including **race and ethnicity, religion, age** and **gender** could lead to an increased risk of harm to individuals.

1.38 Evidence has shown that many terrorist groups or terrorists follow an ideology that targets a specific group. For far-right terrorist groups, this is often minority ethnic groups.[110] These minority ethnic groups may become victims of both online and offline abuse due to a terrorist group's conspiracy theories and racist rhetoric which they spread online.[111] Similarly, proscribed organisations are more likely to target specific socio-cultural groups for radicalisation and recruitment, based on their ideologies and goals. For example, Islamist extremist groups are more likely to attempt to recruit and draw support from Muslim men; while extremist right-wing groups are more likely to focus on recruiting young white men.

1.39 Data from the Online Experiences Tracker shows that males (10%) are more likely than females (7%) to have experienced or encountered content that encouraged extremism, radicalisation or terrorism in the past four weeks. Those from minority ethnic groups (13%) are more likely than the average (8%) to encounter this content and those aged 18 to 24 (13%) are also more likely than average to do so (8%).[112]

1.40 Children may be at increased risk of radicalisation due to the amount of time many of them spend online. Evidence indicates that this may have been exacerbated by social isolation during the COVID-19 pandemic.[113] While current evidence is inconclusive regarding whether children are relatively more likely than adults to encounter terrorism content online, the involvement of children in terrorism cases in the UK is an important emerging trend. In the year ending March 2022, people under the age of 20 represented nearly 3 in 5 (59%) of the 6,393 referrals to Prevent, a Government programme to help safeguard people from becoming terrorists or supporting terrorism.[114] The International Centre for the Study of Radicalisation has also described the subjects of their emerging research into children's involvement in terrorism as 'innovators and amplifiers ', highlighting four cases in which a child was convicted for attempting or managing to establish an online terror network between 2016 and 2023.[115]

1.41 Users who actively seek out terrorist content are also more likely to encounter terrorist content online than those who do not.[116] There is limited evidence on child terrorism activity in the UK and we are looking to build our evidence base regarding how the online environment may enable children to either be radicalised by others or undertake their radicalisation journey.[117]

1.42 In the Register of Risks chapter 'Child sexual exploitation and abuse' we discuss violent online groups which target children aged 8 to 17, particularly those identifying as LGBTQ+,

---

[110] Foreign Policy Magazine (Ware, J. and Clarke, C.P.), 2022. How Far-Right Terrorists Choose Their Enemies [accessed 3 July 2023].
[111] Ofcom, 2022. The Buffalo attack: Implications for online safety. [accessed 29 June 2023].
[112] Ofcom, 2024. Online Experiences Tracker - Wave 6. [accessed 11 November 2024].
[113] Europol, 2022. European Union Situation and Trend Report 2022 [accessed 12 June 2024].
[114] Home Office, 2023. Individuals referred to and supported through the Prevent Programme, April 2021 to March 2022. [accessed 12 June 2024].
[115] Rose, H. and Vale, G., International Centre for the Study of Radicalisation, 2013. Childhood Innocence?: Mapping Trends in Teenage Terrorism Offenders [accessed 27 June 2024].
[116] Schumann, S., Clemmow, C., Rottweiler, B. and Gill, P., 2024. Distinct patterns of incidental exposure to and active selection of radicalizing information indicate varying levels of support for violent extremism, *PLoS ONE* 19(2). [accessed 5 June 2024].
[117] We recommend readers refer to the Protection of Children's Register of Risks to understand the role that legal content can play in radicalising children online.

with mental health conditions or in ethnic minority groups, and manipulate or threaten them into performing various acts which can include mass shootings and spreading far-right terrorist ideology.[118]

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles and fake user profiles

1.43 Terrorists can exploit services that allow users to create an anonymous user profile or allow users to create fake user profiles by altering their usernames.

1.44 Research indicates that anonymous user profiles can empower potential perpetrators by emboldening them to speak and act more radically.[119]

1.45 Users can evade disciplinary action taken by services by creating a different user profile with an altered username. A report from the ISD identified common tactics used among creators of banned accounts, including altering usernames to indicate new versions of previously banned account. Creators then inform their previous followers that they have returned by stating this in their user profile or when commenting on other videos. ISD provides an example of a creator posting, multiple times, a video of a misogynist who killed six people in Isla Vista, California, in 2014.[120]

User networking

User connections and user tagging

1.46 How users can find or encounter like-minded individuals on services is essential in encouraging and promoting terrorism content and activity. Our evidence points to common functionalities such as content or user tagging, hyperlinking, and connecting to a large mass of individuals as being fundamental to this.

1.47 Research by the Global Network on Extremism and Technology (GNET), an academic and research initiative, points to pro-ISIS[121] accounts on mainstream services which have used the "*user tagging function on posts to tag or link to nearly 100 similar accounts".*[122] The research says that the mass-tagging tactic helps amplify terrorism content and allows new audiences to find other pro-ISIS accounts. The research makes the point that although these "*tagging and network methods help terrorism content to spread on mainstream services, they can also be used to identify other ISIS supporter accounts*".[123]

1.48 GNET research also talks about pro-ISIS accounts following, or connecting with, thousands of individual users who can then view, save, and share the extremist content. The research

---

[118] Global Network on Extremism and Technology (Argentino, M., Barrett, G. and Tyler, M. B.), 2024. 764: The Intersection of Terrorism, Violent Extremism and Child Sexual Exploitation [accessed 02 October 2024].
[119] Koehler, D., 2014. The Radical Online: Individual radicalisation processes and the role of the internet, *Journal for Deradicalisation*, Winter 2014/15 (1). [accessed 4 July 2023].
[120] Institute for Strategic Dialogue (O'Connor, C.) 2021. Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. [accessed 4 July 2023].
[121] ISIS (Islamic State of Iraq and Syria), also known as ISIL (Islamic State of Iraq and the Levant), is a proscribed terrorist organisation based in Syria. Source: RAND Corporation, n.d. The Islamic State (Terrorist Organisation). [accessed 23 August 2023].
[122] Global Network on Extremism and Technology (McDonald, B.), 2022. Extremists are Seeping Back into the Mainstream: Algorithmic Detection and Evasion Tactics on Social Media Platforms. [accessed 4 July 2023].
[123] Global Network on Extremism and Technology (GNET) (Basra, R.), 2022. The Future is Now: The Use of 3D-Printed Guns by Extremists and Terrorists. [accessed 22 October 2024].

says that although official pages or channels spreading pro-ISIS content may be removed, a large network of individual users can quickly amass thousands of connections. GNET gives the example of Facebook, where it found that many pro-ISIS accounts had "*maxed out the 5,000 'friend connections' allowed by the platform"*. According to the GNET, the thousands of 'friend connections' by *"pro-ISIS accounts on this platform are large enough to match or even exceed the audiences of many official ISIS channels found on encrypted alt-tech*[124] *platforms"*.[125]

User communication

Livestreaming

1.49 There are many examples of livestreaming being used to promote terrorism. For instance, in 2013, an attack in Kenya was live-tweeted;126 the attack in Buffalo, New York was livestreamed (and versions of the footage were disseminated on multiple online services);127 and an attack in Christchurch, New Zealand was livestreamed.128 129 Between 2016 and 2022, four other terrorist attacks were livestreamed and another failed. Most of the attacks were perpetrated by individuals who appear to have been motivated by far-right ideology.130 The use of livestreaming remains a persistent feature of far-right lone attackers, many of whom directly reference and copy aspects of previous attacks.131 The risks associated with such content are linked to the ability of the content to go viral and motivate others to carry out such attacks in a similar manner.132 This is mostly linked to the ability for the content to be recorded, the details of which can be found in the online payments and crowdfunding section of the chapter.

Live audio

1.50 Live audio, as well as messaging functionalities, is also used on gaming-related content to disseminate terrorism content and radicalise and recruit others to their cause. A paper from the United Nations Office of Counter-Terrorism says this is to generate attention and increase the familiarity and attractiveness of their propaganda in the eyes of the target audience. Moderation in online games is often focused on profanity and/or in-game

---

[124] "*Alt-tech platforms are a group of alternatives to mainstream web and social networking platforms that generally have a smaller staff with fewer capabilities, resulting in low moderation that allows users to post content more freely".* Source: The Counterterrorism Group, (Finnerty, C and Grelicha, K.), 2022. The use of social media and alt-tech platforms by threat actors. [accessed 19 September 2023].

[125] Global Network on Extremism and Technology (GNET) (Basra, R.), 2022.

[126] *"Al-Shabaab, the al-Qaeda affiliate based in East Africa, live-tweeted an attack on the Westgate mall in Nairobi, Kenya, explaining and justifying its actions as it killed 67 people"*. Source: Cronin, A. K., 2020. Power to the People – How Open Technological Innovation is Arming Tomorrow's Terrorists. Oxford University Press, p.189.

[127] Ofcom, 2022. The Buffalo attack: Implications for online safety. [accessed 29 June 2023].

[128] *"A gunman went live on a social media service before he shot and killed 51 people at local mosques. In the same year, a gunman in Germany also livestreamed his attack on a social media service".* Source: Brooks, A and Matromarino, J.P., 2022. Extremists exploit gaming networks and social media to recruit and radicalize, Wbur, 19 May. [accessed 19 September 2023].

[129] In 2016, a man in France used a social media live feature to broadcast his justification for killing two police officers whilst holding a child hostage and pledging his allegiance to the Islamic State. In 2019 a gunman reportedly livestreamed himself through his channel on a social media service, attacking a synagogue and a kebab shop in Halle, Germany. In 2020, an attacker livestreamed himself carrying out an attack in a mall in Glendale, Arizona. Source: Ofcom, 2022. The Buffalo attack: Implications for online safety. [accessed 29 June 2023].

[130] Ofcom, 2022. The Buffalo attack: Implications for online safety. [accessed 29 June 2023].

[131] Andrews, S. 2023. The 'First Person Shooter' Perspective: A Different View on First Person Shooters, Gamification, and First Person Terrorist Propaganda, *Games and Culture* 19(1). [accessed 4 July 2023].

[132] Bryden, M., Bahra, P., Cruickshank, P., Macklin, G., Cook, J., Vale, G and Simcox, R., 2019. CTCSENTINEL. [accessed 26 July 2023]; Kupper, J., Christensen, T. K., Wing, D., Hurt, M., Schumacher, M., Meloy, R., 2022. The_Contagion_and_Copycat_Effect_in_Transnational_Far-right_Terrorism_An_Analysis_of_Language_Evidence. *Perspectives on Terrorism* 16(4), 4-26. [accessed 4 July 2023].

behaviours leading to these gaming-related online "*spaces providing extremists with the opportunity to broadcast their messages widely, and relatively undisturbed*".[133]

Direct messaging, group messaging, encrypted messaging, and ephemeral messaging

1.51 Messaging functionalities such as group messaging and direct messaging can allow terrorists to share content in a low-friction way with large numbers of like-minded people. Encryption and ephemerality make messaging particularly attractive to terrorist actors as they can reduce the chance of detection.

1.52 A paper by Brian Fishman indicates that closed groups in private messaging services, restricted spaces on social media services and discussion forums and chat rooms, all help to separate in-group participants from outsiders.[134]

1.53 A Tech Against Terrorism paper says that terrorists and violent extremists consider multiple functionalities when choosing a service. Although end-to-end encryption is important to them, they will also consider the reach of a service, its ease of use, storage capacity, as well as security features such as ephemeral messaging and password protection, privacy and security. The paper goes on to say that services which use end-to-end encryption, and the mentioned features are preferred by terrorist actors.[135]

1.54 Group messaging or 'group chats' in private messaging services with end-to-end encryption are used as beacons by terrorists, acting as a signpost to the content, without the risk of it being removed by services' moderation teams.[136]

Posting content (text, images, videos)

1.55 Evidence shows that services where content can be posted or shared on an open channel of communication can be conducive to the spread of terrorism content. Terrorists have exploited these functionalities to disseminate material to a wide range of interested parties.

1.56 For example, an article describes how, on the 20th anniversary of the September 11 attacks, a coalition of alt-jihadist[137] meme producers ran a competition to see who could create the best meme of the attacks. "*This challenge was shared through a central page on Facebook, coordinated on Telegram, and A/B[138] tested on Discord*".[139] The researchers found that key accounts across services began creating terrorism content using popular internet memes.[140]

1.57 A UN report describes how services "act as an alternative training ground for terrorists". They enable the sharing and dissemination of material such as "detailed instructions, often in easily accessible multimedia format and multiple languages, on topics such as how to join terrorist organisations; how to construct explosives, firearms or other weapons or

---

[133] United Nations Office of Counter-Terrorism, 2022. Examining the Intersection Between Gaming and Violent Extremism. [accessed 4 July 2023].
[134] Fishman, B. 2019. Crossroads: Counter-terrorism and the Internet, Texas National Security Review, 2(2), 82-100. [accessed 4 July 2023].
[135] Tech Against Terrorism, 2021. Terrorist use of E2EE: State of play, misconceptions, and mitigation strategies. [accessed 4 July 2023].
[136] Tech Against Terrorism report, 2021.
[137] *"Alt-jihadists draw on the narratives of the alt-right and far right in Western culture wars while staying on brand with support for staple extremist groups such as Hezbollah, the Houthis, Hamas, the Taliban, Hayat Tahrir al-Sham, al-Qaeda, and the Islamic State."* Source: Ayad, M., 2021. An 'Alt-Jihad' is Rising on Social Media, Wired, 8 December. [accessed 4 July 2023].
[138] *"A/B testing is a way to compare two versions of something to figure out which performs better. While it's most often associated with websites and apps, the method is almost 100 years old and it's one of the simplest forms of a randomized controlled experiment".* Source: Harvard Business Review, 2017. A refresher on A/B testing. [accessed 26 July 2023].
[139] Ayad, M., 2021.
[140] Ayad, M., 2021.

hazardous materials; and how to plan and execute terrorist attacks".[141] The services make it easy for material to be shared among a large group of people, and can also help build a sense of community among individuals in different locations and with different backgrounds, "encouraging the creation of networks for the exchange of instructional and tactical material".[142]

1.58 The Institute for Strategic Dialogue have identified examples of terrorist groups spreading propaganda by adding original audio to videos on social media, including speeches from members of ISIS, which can be reused even if the video is taken down.[143]

Commenting on content

1.59 The ability to leave comments on a post is another way in which terrorism content can be promoted and disseminated. In Gaming and Extremism, The Extreme Right on Discord, the authors discuss several instances in which users shared comments about AWD,[144] *"with some inquiring about how to find AWD's website and others indicating that they would like to join the group"*.[145]

1.60 Research from GNET has found that the use of 'outlinking' (also known as 'hyperlinking'), which is normally facilitated through comments or posts, is an important tactic used by ISIS supporters. The report says that *"the profiles of pro-ISIS accounts on mainstream social media services can serve as a gateway and landing page to direct users to more explicit terrorism content"*.[146]

1.61 There is evidence to suggest that terrorist organisations use bots[147] to share comments on social media services and are also using the services to collect information. This could be to look for potential recruits, monitor news or use online mapping tools to plan attacks. In September 2021, an investigation by ISD identified a digital library or archive of content belonging to ISIS. It was reported to have contained over 90,000 items and to have received an estimated 10,000 unique visitors per month. This material is then added to social media comment pages and spread via bot accounts.[148]

1.62 Evidence shows that pro-ISIS supporters on mainstream social media services use well-known symbols, hand gestures and emojis to indicate affiliation and support for the group.[149] Examples identified by Bellingcat, an independent investigative journalism group, include memes produced in support of the Christchurch attacker and the Charleston mass shooter.[150]

---

[141] United Nations Office on Drugs & Crime, 2012. The use of the Internet for terrorist purposes. [accessed 4 July 2023].
[142] United Nations Office on Drugs & Crime, 2012.
[143] Institute for Strategic Dialogue (Ayad, M.), 2023. CaliphateTok: TikTok continues to host Islamic state propaganda [accessed 14 June 2024].
[144] Atomwaffen Division (AWD), an extreme right-wing group.
[145] Institute for Strategic Dialogue (Gallagher, A., O'Connor, C., Vaux, P., Thomas, E., Davey, J.), 2021. Gaming and Extremism, The Extreme Right on Discord. [accessed 4 July 2023].
[146] Global Network on Extremism and Technology (GNET) (Basra, R.), 2022. The Future is Now: The Use of 3D-Printed Guns by Extremists and Terrorists. [accessed 22 October 2024].
[147] Bots is an umbrella term that refers to a software application or automated tool that has been programmed by a person to carry out a specific or predefined task without any human intervention.
[148] Silva, S., 2020. Islamic State: Giant library of group's online propaganda discovered, *BBC,* 4 September. [accessed 19 September 2023]; Institute for Strategic Dialogue, 2020. Click reveals ISD discovery of huge pro-ISIS online cache. 8 September. [accessed 19 September 2023].
[149] Global Network on Extremism and Technology (GNET) (Basra, R.), 2022.
[150] Bellingcat, (R. Evans), 2021. White Boy Summer, Nazi Memes and the Mainstreaming of White Supremacist Violence. [accessed 4 July 2023].

Transactions and offers

Online payments and crowdfunding

1.63 Terrorist actors have been known to use functionalities such as the ability to raise funds or crowdsource as ways to encourage engagement with terrorist activity or terrorist actors or to finance acts of terrorism. Services accepting online payments can assist perpetrators in raising funds or crowdsourcing more easily.

1.64 A UN report says that how terrorists use the internet to raise and collect funds and resources can be categorised into: "four general categories: direct solicitation, e-commerce, the exploitation of online payment tools and through charitable organisations […] online payment facilities offered through dedicated websites or communications platforms make it easy to transfer funds electronically between parties. Fund transfers are often made by electronic wire transfer, credit card or alternate payment facilities available via services such as PayPal or Skype".[151]

1.65 It continues: "Financial support provided to seemingly legitimate organizations, such as charities, may also be diverted for illicit purposes […] some terrorist organisations have been known to establish shell corporations, disguised as philanthropic undertakings, to solicit online donations. These organisations may claim to support humanitarian goals while donations are used to fund acts of terrorism".[152]

1.66 The GNET reports that terrorists are increasingly exploiting technology to raise money. "Raising money from support networks is not a new method of fundraising […] it's been enabled and amplified by technology". The report continues: "Sales of propaganda, for example, sales of merchandise or crowdfunding for specific issues also presents a revenue stream for terrorism" and that cryptocurrencies such as Bitcoin and Monero are reported to have been used by ISIS.[153]

Content navigation

Hyperlinking

1.67 Hyperlinking allows terrorist groups or actors to disseminate information to a large audience.

1.68 Evidence highlights that many terrorist organisations or terrorist actors use larger social media services as gateways to the content, as they provide a greater reach and larger audience for their content. Content on larger services will often be a hyperlink to a third-party service, or a copy of the content will be shared on the service, while the source material remains on a smaller service. This is often done to avoid content moderation detection. Research into terrorist use of larger sites tracked 11,520 posts linking to 244 separate host sites.[154]

1.69 Although terrorist-operated websites that are not U2U services are out of scope of this chapter, they form an important part of the wider ecosystem of terrorism content available

[151] United Nations Office of Drugs and Crime, 2012. The Use of the Internet for Terrorist Purposes. [accessed 29 June 2023].
[152] United Nations Office on Drugs & Crime report, 2012.
[153] Global Network on Extremism and Technology (GNET) (Davis, J.), 2021. Technology and Terrorist Financing. [accessed 4 July 2023].
[154] Observer Research Foundation (Saltman, E.), 2022. Identifying and Removing Terrorist Content Online: Cross-Platform Solutions. [accessed 4 July 2023].

online as users may use hyperlinks to them on regulated U2U services. This can enable the terrorist group to bring a wider audience to the services they operate.

1.70 There is evidence that those who actively seek out terrorist content are more likely to encounter terrorist content, and search functionalities may make seeking out terrorist content easier. The ease with which some content can be found via search functionality is compounded by a concern that those who are actively searching for such content may be more susceptible to experiencing or causing harm as a result.[155]

Content storage and capture

Screen capturing or recording

1.71 Screen capture and recording functionalities, as well as other functionalities that allow users to store content, can help to enable the dissemination of terrorism content.

1.72 U2U services which enable livestreaming, have been used to broadcast terrorist attacks in real-time. Although content may only be live for a short period, there is evidence to suggest that the content can be recorded and re-shared or forwarded, thereby increasing its virality[156] and the risks of harm to many individuals. This can be done by using the functionality of a service to capture parts of the livestream. Even where such functionality is not available or is not used, a committed user can use third-party software to record footage of an attack to disseminate on services and forums.[157]

1.73 Terrorist organisations will consider the storage capacity of a service along with its reach, usability, and other security features.[158] The Tech Against Terrorism response to Ofcom's Call for Evidence says that content "*produced by designated terrorist entities is particularly prevalent on file-hosting, archiving, pasting, and video-sharing services which act as content stores*".[159]

Content editing

Editing visual media

1.74 The editing of images and video content can help enable terrorism offences, and there is evidence to suggest that terrorism content can be hidden in images which can then be shared. As referenced earlier, this editing of images can also be used in recruitment, particularly the recruitment of younger people.

1.75 The report '*An 'Alt-Jihad' is rising on social media'* collected over 5,000 memes and videos which had been created and shared by 'alt-jihadists'[160] and the digital communities around them, across many services. Approximately a fifth of these pieces of content supported

---

[155] For instance, research exploring the impact of exposure to potentially radicalizing information suggests that individuals who actively seek out terrorism content are at a higher risk of radicalisation. Source: Schuman, S., Clemmow, C., Rottweiler, B., Gill, P. 2024. Distinct patterns of incidental exposure to and active selection of radicalizing information indicate varying levels of support for violent extremism. [accessed 30 August 2024].

[156] Ofcom, 2022. The Buffalo attack: Implications for online safety. [accessed 29 June 2023].

[157] Ofcom, 2022. The Buffalo attack: Implications for online safety. [accessed 29 June 2023].

[158] Fishman, B. 2019. Crossroads: Counter-terrorism and the Internet Texas National Security Review, 2 (2). [accessed 4 July 2023].

[159] Tech against Terrorism response to Ofcom's Call for Evidence dated 2022.

[160] "*ISD has identified a networked community of 'alternative' support for groups like Al Qaeda and the Islamic State, blending the aesthetics of 'chan culture' the alt-right, and extremist groups. This community of supporters, whom ISD researchers are referring to as 'alt-jihadists', specialises in producing and disseminating Salafi-jihadist content using familiar 'chan culture' and alt-right meme characters such as Pepe the Frog."* Source: Institute of Strategic Dialogue, 2022. Looking Beyond the Traditional threat: Alt-Jihadism. [Accessed August 23 2023].

militant groups, including Hamas and jihadist organisations.[161] There is also evidence that steganography, the hiding of messages in images, is in widespread use by terrorist organisations.[162]

Recommender systems

Content recommender systems

1.76 Although terrorism content is often not permitted under the terms and conditions of online services, there is evidence to suggest that content recommender systems can increase the risk of exposure to it if it is present on a service. Such content could remain on a service for a variety of reasons such as a classification error by automated content moderation systems, inaccurate judgement by a human moderator, or while flagged content is awaiting review.

1.77 While design choices might vary by service, content recommender systems are commonly designed to optimise user engagement. Where users might seek out and engage with terrorism content, content recommender systems may suggest similar content yet to be detected and taken down. Provided there is terrorism content available for recommendation and sufficient user engagement with that content, recommender systems may disseminate that content, increasing the risk of users being exposed to it.

1.78 If the user is expressing interest through active engagement (liking, sharing, and commenting) with terrorism content, there is a risk of a 'filter bubble' forming (an echo chamber of thematically homogenous content). In more extreme cases, a rabbit hole may form (increasing in thematic intensity) for those users most inclined to engage with terrorism content. This is relevant for terrorism content; a paper from RUSI (Royal United States Institute), an independent think-tank specialising in defence and security research, found that recommender systems can prioritise 'extreme content'[163] when users have expressed some form of implicit or explicit interest in similar content. This can include regularly viewing content related to 'extreme content', or explicitly engaging with it by liking and sharing it. [164] Therefore, if such content is not adequately removed by content moderation systems, recommender systems could end up promoting it to users who may have already engaged with similar content.

Risk factors: Business models and commercial profiles

Revenue model

1.79 There is limited evidence on how the different revenue models may affect the risks of harm, related to terrorism, so we have not sought to assess which models are relatively high-risk or compare risk across different revenue models.

1.80 Nevertheless, the evidence below suggests that advertising models may sometimes reduce the risks of harm related to terrorism if advertisers put pressure on services. The initiatives taken by advertisers that have set out the "*goal of eliminating harmful online content and*

[161] Ayad, M., 2021. An 'Alt-Jihad' is Rising on Social Media, Wired, 8 December. [accessed 4 July 2023].
[162] United Nations Office of Drugs and Crime, 2012. The Use of the Internet for Terrorist Purposes. [accessed 29 June 2023].
[163] The term 'extreme' is based on the Holbrook's Extremist Media Index definition. Source: Holbrook, D., 2015. Designing and Applying an 'Extremist Media Index'. *Perspectives on Terrorism,* 9(5). [accessed 20 September 2023].
[164] RUSI (Reed, A., Whittaker, J., Votta, F. and Looney, S.), 2019. Radical Filter Bubbles: Social Media Personalisation Algorithms and Extremist Content. [accessed 4 July 2023].

ensuring that bad actors have no access to advertiser funding"[165] (Global Alliance for Responsible Media) show how advertisers have a role in protecting individuals against harm such as terrorism.

1.81 Indeed, a European Commission paper shows that "social networks and media-sharing platforms, whose business model is often based on advertising, have faced intense public criticism when terrorism content or other illegal content has been reported on their services. This has in some cases triggered direct revenue loss, following a major backlash from certain advertisers, and has led to user distrust".[166]

Commercial profile

Capacity and maturity

1.82 Evidence suggests that low-capacity or early-stage U2U services can present risks of harm for terrorism content as they can be targeted by perpetrators. This is due to their limited technical and financial resources to moderate content, compared to more highly moderated mainstream services.[167]

1.83 Terrorist actors can move from larger, mainstream social media services to smaller services. The size of a service, as it relates to the size of its user base, is discussed in the section 'User Base Size'. However, our research shows that sometimes services with a small user base can have fewer resources to moderate content. Indeed, according to Tech Against Terrorism, this migration to smaller services is, in part, a response to larger services increasing their content moderation functions in recent years. This is supported by evidence showing an ongoing migration of terrorist actors to emerging video-sharing services that have more permissive terms and conditions and do not significantly invest in content moderation systems.[168]

1.84 Different research conducted by Tech Against Terrorism concludes that smaller and newer services are most at risk of exploitation,[169] as terrorists and violent extremists such as ISIS may use them.[170] This includes micro-services that may be run by a single individual. This is largely due to targeting and a lack of technical and financial resources for effective moderation.[171]

---

[165] World Federation of Advertisers, 2020. Marketing leaders take action on harmful online content. [accessed 4 July 2023].
[166] Section 2.3.1 of the European Commission's Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online*: "Hosting service providers are abused for the dissemination of terrorist content online affecting the business models and users' trust in the digital single market"* Source: European Commission, 2018. Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online. [accessed 4 July 2023].
[167] This is different from our analysis of user base size set out earlier. Low capacity is related to size in terms of number of employees and/or revenue, which may increase risk due to ability to moderate, while risks from large services (i.e. with large user base) are related to reach.
[168] Tech Against Terrorism, 2021. Trends in Terrorist and Violent Extremist Use of the Internet. [accessed 29 June 2023].
[169] Tech Against Terrorism, 2019. Analysis: ISIS Use of smaller platforms and the DWeb to share terrorist content – April 2019. [accessed 4 July 2023].
[170] Service size as it relates to the size of its user base is discussed under user base size. However, our research shows that sometimes services with a small user base can have fewer technical and financial resources.
[171] During a panel event hosted for its UK launch, Tech Against Terrorism explained that *"terrorists exploit an overlapping ecosystem of services, not just the big platforms like Facebook and Twitter but also the smaller services.*" The initiative expressed the concern that smaller technology companies are at risk of being exploited by terrorist groups when disseminating propaganda but often do not have the scale or resources to tackle terrorism content or to comply with legal requirements. Source: Tech Against Terrorism, 2017. UK Launch of Tech Against Terrorism at Chatham House. [accessed 4 July 2023].

1.85 According to a report from Europol, while action by industry and law enforcement has reduced terrorist abuse of mainstream service providers, *"similar progress has yet to be made with start-up social media and companies with limited resources*", which are being targeted due to their lower capacity for, and focus on, effectively moderating content.[172]

[172] Europol, 2018. European Union Terrorism situation and trend report. [accessed 4 July 2023].

2. Child Sexual Exploitation and Abuse (CSEA)

Warning: This chapter contains content that may be upsetting or distressing in relation to CSEA.

Introduction

2.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to the Child Sexual Exploitation and Abuse (CSEA)[173] offences listed under 'Relevant offences'; and
- The use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

2.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

2.3 CSEA involves forcing or enticing a child or young person to take part in sexual activities of varying types and severity whether or not the child is aware of what is happening. The activities may involve physical contact or non-contact activities. Sexual abuse can take place online, and technology can also be used to facilitate offline abuse.[174]

2.4 Schedule 6 of the Online Safety Act (the Act) sets out priority offences relating to CSEA.[175] Our assessment of the relevant risks is split into two broad categories: grooming, and child sexual abuse material (CSAM). **We recognise that CSEA offences are complex and that some instances of exploitation or abuse may involve the commission of a number of these specific offences, across both elements of grooming and CSAM.** This is particularly true for offences which involve: sexual exploitation, blackmail, or extortion; coercion, production and sharing of self-generated indecent images (SGII); and child-on-child sexual exploitation and abuse. Ofcom recognises that these types of CSEA can be closely related and can happen either in isolation, in parallel or sequentially. These cross-cutting issues are explored in more detail in 'How CSEA offences manifest online'. This introduction addresses both grooming and CSAM, highlighting relevant offences and how they manifest online. Following this are specific subsections on grooming and CSAM, which share similar

[173] We will set out and describe the evidence we have available about Child Sexual Exploitation and Abuse, which will require us to refer regularly to terms that are long and which, when replaced with acronyms, make reading the document a swifter and neater experience. As is common practice, each term is written out in full upon its first usage, followed by its acronym in parenthesis, which will be used on each occasion afterwards. Our use of acronyms does not detract from the significance and gravity of these terms and in no way reflects any intent to diminish their seriousness.
[174] HM Government, 2023. Working Together to Safeguard Children. [accessed 24 September 2024].
[175] CSEA offences vary in relation to the age, or reasonably perceived age, of the victim; some offences may relate to children under the age of 16 while others apply to all children under the age of 18. In addition, other offences consider factors such as the power imbalance between the perpetrator and the child, namely those in a position of trust. For a full overview of these offences please see the Illegal Content Judgements Guidance (ICJG).

structures and overlapping content. However, the evidence in each section focuses on these harms individually. The subsections explore in greater detail:

a) How these specific harms manifest online
b) Evidence of risk factors in user-to-user (U2U) services
c) Specific risk factors related to service types
d) Risk factors associated with user bases, such as size and demographics
e) Risk factors related to platform functionalities and recommender systems
f) Risk factors linked to business models and commercial strategies

2.5 Ofcom recognises that each victim and survivor has a unique experience, both at the time of and after they are offended against. Victims and survivors of CSEA often take longer than five years (with an average of 17 years) to disclose the abuse.[176] We highlight the trauma and impact of these experiences to help services understand the risk of the harm. These insights come from research undertaken with individuals with lived experience of the harm and are not intended to speak on their behalf.

2.6 **While this chapter addresses CSEA offences, we acknowledge that CSEA can manifest in different and complex ways. For an understanding of wider risks to children from illegal content, refer also to Human trafficking chapter (the evidence in this chapter is particularly relevant to the sexual exploitation of children[177] [178]). There is also a cross-over** between children who experience CSEA and those who are targeted for criminal exploitation and other human trafficking offences. For more information, refer to the Human Trafficking, Firearms, knives and other weapons, Suicide, and Terrorism chapters.

Relevant offences

2.7 This section provides an overview to some of the relevant offences which we will consider in this chapter.

2.8 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding CSEA, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 6 of the Act.

---

[176] Cited in Halvorsen, J. E. & Tvedt Solberg, E. & Hjelen Stige, S., 2020. "To say it out loud is to kill your own childhood." – An exploration of the first person perspective of barriers to disclosing child sexual abuse, *Children and Youth Services Review,* Elsevier, 113. [accessed 22 September 2023].

[177] Child sexual exploitation is a form of child sexual abuse that occurs where an individual or group takes advantage of an imbalance of power to coerce, manipulate or deceive a child or young person under the age of 18 into sexual activity (a) in exchange for something the victim needs or wants, and/or (b) for the financial advantage or increased status of the perpetrator or facilitator. The victim may have been sexually exploited, even if the sexual activity appears consensual. Child sexual exploitation does not always involve physical contact; it can also occur through the use of technology. This is explained in greater detail in the **Human trafficking chapter**. See also, Department for Education, 2017. Child sexual exploitation. [accessed 22 September 2023].

[178] An offence under the following provisions of the Sexual Offences Act 2003: section 47 (paying for sexual services of a child); section 48 (causing or inciting sexual exploitation of a child); section 49 (controlling a child in relation to sexual exploitation); section 50 (arranging or facilitating sexual exploitation of a child). An offence under the following provisions of the Sexual Offences (NI) Order 2008 (S.I. 2008/1769 (N.I. 2)): article 21 (arranging or facilitating commission of a child sex offence); article 37 (paying for the sexual services of a child); article 38 (causing or inciting child prostitution or pornography); article 39 (controlling a child prostitute or a child involved in pornography); article 40 (arranging or facilitating child prostitution or pornography). An offence under the following provisions of the Protection of Children and Prevention of Sexual Offences (Scotland) Act 2005: section 9 (paying for the sexual services of a child); section 10 (causing or inciting provision by child of sexual services or child pornography); section 11 (controlling a child providing sexual services or involved in pornography); section 12 (arranging or facilitating provision by child of sexual services or child pornography).

2.9 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of the offences listed below (and, in relation to offences in Scotland, being involved art and part in the commission of those offences).

2.10 For the grooming category, we consider the following offences:

- Causing or inciting a child to engage in sexual activity[179]
- Engaging in sexual activity in the presence of a child[180]
- Causing a child to watch a sexual act[181] or to look at a sexual image[182]
- Arranging or facilitating commission of a child sex offence[183]
- Meeting a child following sexual grooming[184] or following certain preliminary contact[185]
- Sexual communication with a child[186] or communicating indecently with a child[187]

2.11 For the CSAM category, we consider the following offences:

- Publishing an obscene article tending to deprave and corrupt others by encouraging them to commit a relevant offence[188]
- Possession of prohibited images of a child[189]
- Possession of a paedophile manual[190]
- Taking/making, distribution, possession and publication of indecent photograph or pseudo-photograph of a child[191]
- offences relating to taking/making, distribution, possession or publication of indecent photographs of children in Scotland[192]
- offences relating to taking, distribution, possession, or publication of indecent photographs of children in Northern Ireland[193]

---

[179] For children under the age of 13: section 8 of the Sexual Offences Act 2003, article 15 of the Sexual Offences (NI) Order 2008 (S.I. 2008/1769 (N.I. 2)), and section 21 of the Sexual Offences (Scotland) Act 2009. For children over the age of 13: Section 10 of the Sexual Offences Act 2003, Article 17 of the Sexual Offences (NI) Order 2008 (S.I. 2008/1769 (N.I. 2)), section 31 of the Sexual Offences (Scotland) Act 2009, and section 54 of the Sexual Offences (Scotland) Act 2009.

[180] Section 11 of the Sexual Offences Act 2003 and Article 18 of the Sexual Offences (NI) Order 2008 (S.I. 2008/1769 (N.I. 2)).

[181] Section 12 of the Sexual Offences Act 2003 and Article 19 of the Sexual Offences (NI) Order 2008 (S.I. 2008/1769 (N.I. 2)).

[182] section 23 and 33 of the Sexual Offences (Scotland) Act 2009.

[183] Section 14 of the Sexual Offences Act 2003 and Article 21 of the Sexual Offences (Northern Ireland) Order 2008 (S.I. 2008/1769 (N.I.2)).

[184] Section 15 of the Sexual Offences Act 2003 and Article 22 of the Sexual Offences (NI) Order 2008 (S.I. 2008/1769 (N.I. 2)).

[185] Section 1 of the Protection of Children and Prevention of Sexual Offences (Scotland) Act 2005.

[186] Section 15A of the Sexual Offences Act 2003, article 22A of the Sexual Offences (NI) Order 2008 (S.I. 2008/1769 (N.I. 2)).

[187] sections 24 and 34 of the Sexual Offences (Scotland) Act 2009.

[188] Section 2 of the Obscene Publications Act 1959. For present purposes, the relevant offences are those listed under paragraphs 2, 4, 5, 7, and 8 of Schedule 6 to the OSA.

[189] Section 62 of the Coroners and Justice Act 2009.

[190] Section 69 of the Serious Crime Act 2015.

[191] Section 1 of the Protection of Children Act 1978 and section 160 of the Criminal Justice Act 1988.

[192] Sections 52(1)(b); section 52(1)(d); section 52A of the Civic Government (Scotland) Act 1982.

[193] Article 3 of the Protection of Children (NI) Order 1978 (S.I. 1978/1047 (N.I. 17)).

2.12 For more information on these offences and how services can determine if content amounts to illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

How CSEA offences manifest online

2.13 This section is an overview which looks at how CSEA offences manifest online, and the potential risks of harm to individuals.

Scale of offending and scale of reporting

2.14 While it is difficult to know the true scale of CSEA due to the complexity of the reporting process for victims and survivors and the hidden nature of these crimes, a range of sources provide insight into the scale and severity of the threat. The Crime Survey for England and Wales (CSEW) estimated that in 2019, 7.5% of adults aged 18 to 74 had experienced sexual abuse before the age of 16 – approximately 3.1 million people.[194] In 2022/23, over 105,000 child sexual abuse offences were recorded by police in England and Wales.[195] The UK National Crime Agency estimates that there are between 710,000 and 840,000 UK-based adult offenders who pose varying degrees of risk to children, equivalent to 1.3% to 1.6% of the UK adult population.[196]

2.15 For many victims and survivors, the abuse begins when they are young children.[197] The CSEW estimates that, of those who had experienced sexual abuse before the age of 16, 48% (nearly half) of victims and survivors, the sexual abuse had started or occurred before the age of 11.[198]

2.16 Police-recorded CSEA figures likely underestimate the true scale of CSEA due to underreporting; the Independent Inquiry into Child Sexual Abuse (IICSA) found that 67% (more than two-thirds) of victims and survivors who took part in the study did not disclose their sexual abuse at the time.[199] Similarly, the CSEW found that 76% (more than 3 in 4) of adults who experienced rape or assault by penetration before the age of 16 did not tell anyone at the time, and only 7% reported it to the police at the time.[200]

---

[194] ONS, 2020. Child sexual abuse in England and Wales: year ending March 2019. [accessed 31 July 2023]. The IICSA also reported that research indicates that one in six girls and one in twenty boys are sexually abused before the age of 16Source: IICSA, 2022. : IICSA, 2022. The Report of the Independent Inquiry into Child Sexual Abuse. [accessed 2 October 2023].

[195] CSA Centre (Karsna, K. and Bromley, P.), 2023. Child sexual abuse in 2022/23: Trends in official data. [accessed 11 June 2024].

[196] National Strategic Assessment of Serious and Organised Crime 2024. Child Sexual Abuse - National Crime Agency [accessed 20 September 2024]

[197] The IICSA found that 79% of the participants in the Truth Project were aged 11 or under when they were first sexually abused: IICSA, 2022. The Report of the Independent Inquiry into Child Sexual Abuse. [accessed 31 July 2023]; ONS, 2020. Child sexual abuse in England and Wales: year ending March 2019. [accessed 31 July 2023].

[198] The IICSA found that 79% of the participants in the Truth Project were aged 11 or under when they were first sexually abused: IICSA, 2022. The Report of the Independent Inquiry into Child Sexual Abuse. [accessed 31 July 2023]; ONS, 2020. Child sexual abuse in England and Wales: year ending March 2019. [accessed 31 July 2023].

[199] 5,862 victims and survivors participated in the Truth Project: The IICSA, 2022. The Report of the Independent Inquiry into Child Sexual Abuse [accessed: 2 October 2023].

[200] ONS, 2020. Child sexual abuse in England and Wales: year ending March 2019. [accessed 31 July 2023]

2.17 The scale of online CSEA has been increasing year-on-year, likely further exacerbated by factors such as the COVID-19 lockdown, more time spent online, and efforts to reduce isolation through virtual connections.[201]

2.18 Perpetrators of CSEA offences can come from all age groups and demographics, and the evidence shows there are a range of motivations behind committing such offences. Many perpetrators are motivated by a sexual interest in children; other motivations include curiosity and thrill-seeking.[202] Not all perpetrators have an initial sexual interest in children; research suggests that CSA offending may start during times of stress, difficulties with personal relationships, loneliness, alcohol or drug addiction or depression,[203] and be triggered by the availability of CSAM and access to children online.[204] Some perpetrators have other motivations, including financial gain.[205] Perpetrators may be known or unknown to the victim and survivors.

2.19 While the main focus of this chapter is on adult perpetrators, there are cases where children themselves engage in CSEA. A report from the Vulnerability Knowledge and Practice Programme (VKPP), based on data collected from 42 police forces, found that over half (52%) of CSEA cases recorded between January and December 2022 were committed by children aged 10 to 17.[206]

2.20 Throughout this chapter, unless clearly specified we have opted to use the word 'perpetrator[207]' to describe an adult who commits any of the relevant offences discussed in this chapter. Labelling a child as a perpetrator places them within an adult framework, equating their actions and intent to those of adult perpetrators, which would be inappropriate as there are differences in the dynamics of sexual abuse perpetrated by adults and children. However, this does not diminish the harm that victims and survivors may experience because of other children's harmful sexual behaviours.

2.21 Offences related to child-on-child sexual exploitation and abuse[208] also cut across the different types of CSEA. Our analysis of online grooming and CSAM recognises that those committing CSEA offences can themselves be children, particularly in the non-consensual sharing of self-generated indecent imagery (SGII) (see SGII section for more information).

---

[201] EUROPOL, 2020. Exploiting Isolation: Offenders and victims of online child sexual abuse during COVID-19 pandemic. [accessed 31 August 2023].

[202] CSA Centre response to the November 2023 Illegal Harms Consultation.

[203] NatCen (DeMarco, J., Sharrock, S., Crowther, T., and Barnard, M.), 2017. Behaviour and characteristics of perpetrators of online-facilitated child sexual abuse and exploitation: A Rapid Evidence Assessment Final Report. [accessed 22 September 2023]; CSA Centre response to the November 2023 Illegal Harms Consultation.

[204] Babchishin, K., Hanson, R. & VanZuylen, H., 2014. Online Child Pornography Offenders are Different: A Meta-analysis of the Characteristics of Online and Offline Sex Offenders Against Children. *Archives of sexual behaviour*. 44.

[205] For example, through sexually coerced financial gains or through so-called 'invite child abuse pyramid' sites which encourage the sharing of CSA sites to increase traffic to their site: IWF, 2023. The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 22 September 2023].

[206] VKPP 2022 National Analysis of Police-Recorded Child Sexual Abuse and Exploitation Crimes Report 2022? [accessed 25th September 2024].

[207] We recognise that it is preferable in certain situations to refer instead to the behaviour rather than the individual (for example, 'person who sexually abused') but to account for the variety of evidence referring specifically to 'perpetrators' and the variety of different offences an individual may have committed, we have decided to use 'perpetrator' for simplicity.

[208] An offence under section 13 of the Sexual Offences Act 2003 (child sex offences committed by children or young persons) or an offence under article 20 of the Sexual Offences (NI) Order 2008 (S.I. 2008/1769 (N.I. 2)) (child sex offences committed by children or young persons).

Cross cutting harms

2.22 This section explores in more detail the intersection between CSAM and grooming, as well as various cross-cutting harms, including financially motivated sexual extortion, and other egregious harms.

How CSAM and grooming offending overlap

2.23 Online grooming and CSAM are two prominent forms of CSEA. Online grooming is when a person builds a relationship with a child to manipulate and/or sexually abuse them.[209] CSAM is material which depicts penetrative sexual activity, non-penetrative sexual activity, or other indecent or prohibited imagery of children;[210] it can include material which contains advice about grooming or abusing a child sexually, or that is an obscene article encouraging the commission of other CSEA offences. For more information, see the grooming and CSAM chapters.

2.24 Evidence suggests the relationship between grooming and CSAM is bidirectional. Grooming can lead to the generation of CSAM because, typically, the objective of online grooming is the generation of CSAM and contact sexual abuse of children.[211] Contact sexual abuse of a child can occur in person or can involve the perpetrator[212] remotely forcing the victim to engage in sexual acts either alone or with others, including penetrative acts; the abuse is often recorded, either by the perpetrator or the child,[213] and is often then further shared online as 'first-generation' CSAM.[214]

2.25 CSAM can also lead to grooming. Evidence suggests viewing CSAM can increase the risk a perpetrator will contact a child - perpetrators of contact child sexual abuse are likely to have accessed CSAM online before offending in person.[215] Finnish-based charity, Protect Children, ran an anonymous survey on the dark web and found that 37% (more than one

---

[209] HM Government, 2021. Tackling Child Sexual Abuse Strategy 2021, page 86. [accessed 10 October 2024].
[210] Crown Prosecution Service, 2020. Indecent and Prohibited Images of Children. [accessed 18 August 2023].
[211] CEOP, year of publication unknown. What is sexual grooming? [accessed 24 September 2024].
[212] Please note that we have opted to use the word 'perpetrator' to describe an individual who commits any of the relevant offences discussed in this chapter. We recognise that it is preferable in certain situations to refer instead to the behaviour rather than the individual (for example, 'person who sexually abused') but to account for the variety of evidence referring specifically to 'perpetrators' and the variety of different offences an individual may have committed, we have decided to use 'perpetrator' for simplicity."
[213] Contact abuse refers to instances where an abuser physically interacts with a child, such as through inappropriate touching or forcing the child to undress. Conversely, non-contact abuse, involves actions that do not require physical contact, such as encouraging the child to view sexual acts. NSPCC, 2024. Protecting children from sexual abuse. [accessed 25 September 2024].
[214] 'First-generation' or 'novel' CSAM refers to material that is newly generated and which has not been previously shared or re-shared; this is explored further in the CSAM section below. The relative volume of first-generation CSAM online compared to known CSAM is hard to determine as measures for the volume of CSAM circulating online necessarily rely on the identification of known CSAM (for example, via hash-matching). However, the dramatic recent increase in cases of financially motivated sextortion (see further in the document for more information), which occur after the creation of new first-generation material, suggests there is a significant volume of 'novel' CSAM circulating online.
[215] IICSA (Senker, S., Scott, M. and Wainwright, L.), 2020. An explorative study on perpetrators of child sexual exploitation convicted alongside others. [accessed 22 September 2023]; Cale, J., Holt, T., Leclerc, B., Singh, S., & Drew, J., 2021. Crime commission processes in child sexual abuse material production and distribution: A systematic review. *Trends and issues in crime and criminal justice,* 617. [accessed 22 September 2023].

third) of surveyed individuals who had viewed CSAM tried to seek direct contact with a child afterwards.[216]

2.26 In addition to this bidirectional relationship, there are also some types of CSEA that cut across both CSAM and grooming, such as livestreaming and financially motivated sexual extortion.

Increasingly, advanced technology has enabled a dramatic rise in the creation of Self-Generated Indecent Imagery (SGII)

2.27 Self-generated indecent imagery (SGII) is CSAM produced by a child, in which a perpetrator does not appear to be present in the image/video. Some SGII is coerced from children via grooming.[217] In recent years, there has been a dramatic increase in the presence of SGII online and it now accounts for the majority of reports (92%) actioned by the Internet Watch Foundation (IWF).[218]

2.28 SGII can be defined as either consensual, non-consensually distributed, or aggravated. Regardless of how the content is made, it is considered as CSAM by law enforcement in terms of an adult offender's possession and distribution of SGII.

- **Consensual SGII** is produced and shared on a consensual basis by children to express themselves online, and/or as part of their exploration of their own sexuality.[219] Research has shown how this presents on social media services, and how it is used to initiate and maintain romantic or peer relationships.[220]
- **Non-consensual SGII** occurs when these images are then shared onwards without consent, sometimes quickly going viral. A recently noted example of this is 'bait-out' pages, where sexual images initially shared in confidence are forwarded and re-shared to much larger networks, for example among peer groups.[221]
- **Aggravated SGII** is imagery of child sexual abuse obtained from a child by a perpetrator, usually as part of the grooming process, using tactics such as deceit, threats, or gifting. It can include images, video, or livestreamed illegal content. The child or children involved may have been manipulated into believing that the sexual abuse is consensual, or they may recognise that they have been forced into producing and sending illegal

---

[216] Suojellaan Lapsia, Protect Children (Insoll, T., Ovaska, A., and Vaaranen-Valkonen, N.), 2021. CSAM Users in the dark web: Protecting children through prevention. [accessed 25 August 2023].

[217] Ofcom is aware of other terms used to describe self-generated indecent imagery (SGII), such as 'youth produced sexual imagery' and colloquial terms such as 'sexting' or 'sharing nudes'. Ofcom has used the term SGII in this statement as the one currently most used within the online child protection system, acknowledging that there is work ongoing to consider alternative descriptors.

[218] The IWF found that of the 275,652 webpages it acted on during 2023, 92% were assessed as containing SGII: IWF, 2023 'Self-generated' child sexual abuse [accessed 24 September 2024].

[219] It is crucial for services to note that consensual SGII is typically created by children, and sent to other children, and although the image itself must be treated as illegal and removed and reported to the appropriate designated body (See ICJG - Volume 5, Chapter 26), it is important that services are aware of the nuances associated with this type of behaviour. There are contextual factors in the creation and uploading of the image that require a wider societal response including adequate policing response and education.

[220] Brook and CEOP (McGeeney, E. and Hanson, E.) 2017. Digital Romance: A research project exploring young people's use of technology in their romantic relationships and love lives.; A UK national survey of 14,944 children and young people found that 17% of 15 to 17-year-olds had shared a sexual image: Internet Matters (Katz, A. and El Asam, A.), 2020. Look at Me: Teens, sexting and risks. [accessed 18 August 2023]; Thorn, 2022. Online Grooming: Examining risky encounters amid everyday digital socialization. Findings from 2021 qualitative and quantitative research among 9-17-year-olds.

[221] Revealing Reality, 2023. Anti-social Media: The violent, sexual and illegal content children are viewing on one of their most popular apps. [accessed 31 August 2023].

images of themselves and/or others. Some perpetrators will coerce children into producing SGII and use the SGII to blackmail children into sharing further SGII, sometimes inciting the child to abuse friends and siblings by threatening to publish the SGII online or send the images to friends and family. Some perpetrators also use SGII obtained through deceit and coercion, using the same form of threats, to blackmail children for financial gain. This is known as financially motivated sexual extortion (FMSE, or 'sextortion').

Financially motivated sexual extortion

2.29 There is a growing trend of SGII being used as a method to blackmail children for financial payments, known as financially motivated sexual extortion (FMSE), or 'sextortion'.[222] These cases often, though not exclusively, follow a process similar to grooming, with many enabling factors for FMSE mirroring those of grooming. This includes the ability to identify and contact children online, as well as to see their wider online friends lists, networks and groups.

2.30 Perpetrators' motivations for committing grooming offences of this nature are typically financially driven; in 2022, reports received by the US National Center for Missing & Exploited Children[223] (NCMEC) found that in 79% of identified cases involving sexual extortion the perpetrators were seeking money. The number of cases reported to NCMEC globally also more than doubled in 2023, rising to 26,718 compared to 10,731 in 2022. All age groups and genders have been targeted, but 91% of victims of FMSE dealt with by the Internet Watch Foundation in 2023 were male, and a large proportion of cases have involved male victims aged between 14 and 18. FMSE can impact negatively on victims' and survivors' mental wellbeing, including feelings of shame, guilt, anxiety, fear, and thoughts of self-harm and suicide.[224] At least 27 deaths by suicide have been linked to FMSE in the USA, and there are also known instances of suicides in the UK.[225] [226]

2.31 Perpetrators frequently coerce children into producing SGII using fake account profiles that they believe will appeal to the child victims. Child victims and survivors have reported being contacted by unfamiliar online accounts that appear to belong to other children or young people. The perpetrator typically engages the victim in sexually explicit communication, which may start with them sharing an indecent image. Victims and survivors report being manipulated or pressured into taking nude or semi-nude photos or videos of themselves.[227]

2.32 Images of victims can also be fabricated by the perpetrator, including hacked, digitally manipulated or AI-generated images. The perpetrator typically then threatens to publish the images online and/or send them to the victim's friends and family, and demands payment, for example in the form of money, vouchers or pre-paid gift cards.[228]

---

[222] NCA, 2024. NCA issues urgent warning about 'sextortion' [accessed 24 October 2024]

[223] A US-based non-for-profit organisation which works to help find missing children, reduce child sexual exploitation, and prevent child victimisation.

[224] NSPCC, 2024. Young people's experiences of online sexual extortion or 'sextortion'. [accessed 26 September 2024]

[225] BBC (Tidy, J.), 2024. Dead in 6 hours: How Nigerian sextortion scammers targeted my son, 09 June. [accessed 24 October 2024].

[226] Brooks, L. and Milmo, D. 2024. National Crime Agency threatens extraditions over rise in sextortion cases, The Guardian, 22 August. [accessed 24 October 2024]

[227] NCA, 2024. NCA issues urgent warning about 'sextortion'. [accessed 24 October 2024]

[228] NCA, 2024.

Other egregious harms

2.33 Another serious and cross cutting form of harm to children, facilitated by grooming and extortion involves sadistic interest groups that deliberately target children to threaten, blackmail and extort them into recording or live-streaming acts of extreme self-harm including suicide, as well as animal torture, and to create CSAM and other extreme abuse content.[229] Gaining access to these groups can require members to livestream or upload new footage depicting extreme abuse of children. This footage is then circulated among group members to further blackmail and control victims, perpetuating the cycle of abuse.[230] These groups are reported to operate on gaming platforms, social media sites and mobile applications popular with children. Initial contact is often made through direct messaging before moving to platforms with video enabled features, where conversation can quickly become sexualised or violent.[231] These groups have used threats to SWAT[232] or DOX[233] children, as well as threatening to share images of the victims online and with their friends and family if they do not comply.[234]

2.34 The groups are reported to target children aged between 8 and 17, with a particular focus on LGBTQIA+ individuals, ethnic minorities, and children with a variety of mental health issues, such as depression and suicidal ideation.[235] Victims, including some who are still children, have also been reported to have been recruited to perpetrate similar acts against other children.[236]

2.35 The motivations of these groups appear to be varied and sadistic. The US Federal Bureau of Investigation assesses that the groups' primary motivations include gaining notoriety and status by forcing and threatening children into increasingly extreme acts, while the Royal Canadian Mounted Police assessed the key motivators to be to spread their ideology, gain notoriety, collect extreme violent online content, and rise in status within their groups.[237] Other motivations are reported to include, *inter alia*, nihilism, subversion, an interest in gore and the glorification of mass shooters as well as content related to terrorist and violent extremist content.[238]

---

[229] As well as encouraging self-harm, suicide, and animal torture, these groups also encourage children to victimise others, and stab and 'cut sign' themselves, whereby a victim carves their abuser's name into their body. We have summarised the activities of these groups in this chapter, but the activities they are engaged in cut across various kinds of illegal harm with separate chapters in the Register of Risks. See 'Encouraging and assisting suicide'; 'Encouraging and assisting serious self-harm'; 'Animal cruelty'; 'Human and animal torture' and 'Terrorism' chapters for how these kinds of offences tend to manifest online.

[230] FBI, 2023. Violent Online Groups Extort Minors to Self-Harm and Produce Child Sexual Abuse Material. [accessed 2 November 2024].

[231] RCMP, 2024. RCMP reminds Canadians about violent online groups targeting youth. [accessed 25 October 2024].

[232] The action or practice of making a prank call to police or emergency services in an attempt bring about the dispatch of armed police officers such as a SWAT team to a particular address.

[233] The action of obtaining and publishing personally identifiable information (PII) on the internet, usually for malicious intent.

[234] FBI, 2023. Violent Online Groups Extort Minors to Self-Harm and Produce Child Sexual Abuse Material. [accessed 2 November 2024].

[235] FBI, 2023.

[236] GNET (Argentino, M., Barrett G., and Tyler, M. B.), 2024. 764: The Intersection of Terrorism, Violent Extremism, and Child Sexual Exploitation. [access 2 November 2024].

[237] RCMP, 2024. RCMP reminds Canadians about violent online groups targeting youth. [accessed 25 October 2024]

[238] GNET (Argentino, M., Barrett G., and Tyler, M. B.), 2024. 764: The Intersection of Terrorism, Violent Extremism, and Child Sexual Exploitation. [access 2 November 2024].

2.36 The full scale of offending by such groups is unclear, however multiple groups or offshoots have been associated with this type of offending.[239] While Ofcom considers that the scale is likely to be more limited than for grooming for sexual purposes or for financially motivated sexual extortion, it is likely that offences are under-reported due to the hold that perpetrators can have over their victims, and the extremity and impact on victims of such offending is severe. For more information on the user base risk demographics, see relevant sections of the grooming and CSAM chapters.

Some technologies and functionalities are linked to particular kinds of CSEA

2.37 The role of specific functionalities in enabling CSAM and grooming offences to occur is covered in detail in the subsequent chapters, but the scale and nature of the issue is important to understand, particularly given the speed with which the online landscape can change.

Livestreaming

2.38 Livestreamed CSEA is a widespread problem, both in the UK and globally, where offenders view, direct and comment on the sexual abuse of children in real time.[240] It can involve grooming, blackmail or secretly recording children without their knowledge. Offenders may use screen capture and recording functionalities to create CSAM from livestreamed CSEA. The risk of livestreaming and how it intersects with other functionalities such as messaging, is discussed in the CSAM chapter.

2.39 While livestreamed child sexual abuse can affect any child worldwide, a particular method involves perpetrators from the Global North exploiting economic disparities by paying for the live sexual abuse of children in the Global South. Interpol predicted an increase in this type of exploitation due to the COVID-19 pandemic, as travel restrictions limited some abusers' access to children, and worsened global economic conditions.[241] Indeed, there was a 265% increase in recorded cases of CSEA livestreamed from the Philippines during the quarantine period March to May 2020.[242] Recent estimates have shown that, in 2022, approximately 1 in 100 children in the Philippines were trafficked to produce child sexual exploitation material.[243]

Generative Artificial Intelligence (Gen AI)

2.40 The threat presented by online CSEA is constantly evolving as technology develops, and there is growing evidence that new technologies are being exploited for the commission of CSEA offences.

---

[239] The Royal Canadian Mounted Police reports that one such group targeting children is commonly known as the 764 network (or "the com") but goes by various monikers: RCMP, 2024. RCMP reminds Canadians about violent online groups targeting youth. [accessed 25 October 2024].

[240] WeProtect outlines that the livestreaming of CSEA exists in two main forms: (i) It can be "livestreaming an act of child sexual exploitation and abuse happening offline", or (ii) it can be "one or more children being forced into 'performing' sexual acts in front of a webcam (or camera). This can often be in exchange for payment.": WeProtect. Livestreaming child sexual exploitation and abuse.[accessed 25 August 2023].

[241] Interpol, 2020. Threats and trends Child sexual exploitation and abuse: covid-19 impact. [accessed 22 September 2023].

[242] WeProtect, n.d. Live Streaming Child Exploitation and Abuse. [accessed 22 September 2023].

[243] International Justice Mission and University of Nottingham Rights Lab, 2023. Scale of Harm Research Method, Findings, and Recommendations: Estimating the Prevalence of Trafficking to Produce Child Sexual Exploitation Material in the Philippines. [accessed 12 June 2024].

2.41 There is evidence that generative AI is being used to create increasingly realistic CSAM.[244] This technology generates new abuse images from existing photos of children online, as well as modifying existing CSAM, often making the abuse appear more violent.[245] The BBC reported on 20 girls in Spain, between the ages of 11 and 17, having their fully-clothed imagery manipulated to depict them without their clothes on.[246] The IWF have found material that guides offenders on how to write prompts and train AI to produce more lifelike images.[247]

2.42 Evidence indicates the growing threat of AI-generated CSAM. The University of Stanford's Internet Observatory found a large number of child sexual abuse images in an open-source data set used to train popular AI image generation models.[248]

2.43 As well as generating CSAM, generative-AI chatbots are being used to simulate the offence of sexual communication with a child[249], allowing users to roleplay interaction and abuse of a child, without committing the act of physically abusing a child. There is evidence that some offenders are building AI chatbots companions for the explicit purpose of having "*realistic paedophilic role plays with AI child avatars*".[250]

2.44 Generative-AI is also being used to facilitate grooming offences. This technology can be used to create images and videos which can be uploaded to U2U services and enable approaches to children to seem more genuine, increasing the chance of a response. AI large language models can also be used by perpetrators to quickly learn about topics relevant to their intended victims, helping them to build relationships.

2.45 The risk factors associated with AI-generated CSEA are discussed where relevant throughout the grooming and CSAM sections. For more information on the impacts of AI-generated CSAM, see 'Risks of harm to individuals presented by child sexual abuse material offences.'

Virtual Reality and Augmented Reality

2.46 Virtual reality (VR) and augmented reality (AR) technologies also present risks for child safety. Evidence in understanding the harms that VR and AR technologies present to children is growing. The WeProtect Global Alliance (WeProtect) and the University of Manchester have identified emerging risks, including access to children, CSAM distribution

---

[e] The IWF have reported that most AI CSAM is now realistic enough to be treated as 'real' CSAM, with the most convincing AI CSAM being visually indistinguishable from real CSAM, even for trained IWF analysts: IWF, 2023. How AI is being abused to create child sexual abuse imagery. [accessed 11 June 2024]; In a one-month period between September and October 2023, the IWF were able to scrape 20,254 AI generated images from one dark web forum. Of the 11,108 images assessed using human review, 2,978 images were illegal either under the Protection of Children Act (1978) or the Coroners and Justice Act (2009). While the dark web is not the focus of this analysis, it is reasonable to assume that some of this imagery may eventually be shared to the clear web: Internet Watch Foundation (IWF) response to November 2023 Illegal Harms Consultation.

[245] NCMEC, 2023. What does generative AI mean for CSE? [accessed 12 June 2024].

[246] Hedgecoe, G., 2023. AI-generated naked child images shock Spanish town of Almendralejo. The BBC, 24 September. [accessed 12 June 2024].

[247] IWF, 2023. How AI is being abused to create child sexual abuse imagery. [accessed 11 June 2024]; Active Fence, 2023. How predators are abusing generative AI. [accessed 12 June 2024].

[248] Thiel, D., 2023. Investigation Finds AI Image Generation Models Trained on Child Abuse. Stanford University, 20 December. [accessed 12 June 2024].

[249] Note that, at the time of writing, although this specific activity would not constitute an offence covered by Schedule 6 of the OSA, proposed amendments to the Criminal Justice Bill include a new Clause 26 to make it an offence to use, create or share online digital tools which simulate the offence of sexual communication with a child: House of Commons, 2024. Criminal Justice Bill, As Amended (Amendment Paper).

[250] IWF, 2024. Briefing from the Internet Watch Foundation: Criminal Justice Bill Report Stage. [accessed 15 November 2024]

and simulated abuse.[251] Integrated technologies like haptics, which simulate real world sensations such as movements, vibrations, and force, can make virtual abuse feel more realistic. Further, research by the NSPCC indicates that perpetrators often find new ways to exploit immersive technologies to groom and exploit children.[252] The harm posed by VR and AR to these environments but can easily escalate to offline offences.

2.47 No two experiences of child sexual abuse are the same, and nor are the resulting impacts; evidence highlights the profound and long-lasting effects CSEA can have on victims and survivors. Many describe how childhood sexual abuse has affected different aspects and stages of their lives. CSEA can occur exclusively online, offline or through a combination of both; however, none of these particular forms of CSEA are inherently more or less impactful than another.[253]

2.48 CSEA has been shown to have a negative impact on victims' and survivors' mental health. In one analysis, 88% (nearly 9 in 10) of victims and survivors of sexual abuse reported that it had had a negative impact on their mental health, and almost a fifth said they had attempted suicide.[254] Research also links CSEA to conditions including post-traumatic stress disorder, depression, anxiety, and personality disorders.[255]

2.49 Some victims and survivors describe the negative impacts that CSEA has had on their education, employment prospects, and personal relationships. They describe challenges in their relationships, such as a lack of trust, and difficulties with intimacy or affection due to experiencing sexual abuse. Some also mention becoming overly protective parents or struggling to create and set boundaries.[256]

2.50 In addition to the impact on victims' and survivors' mental health, their physical health can also be affected in the longer term, both due to injuries sustained during, and physical conditions resulting from, their abuse as a child. The CSA Centre outlined how there can be negative physical impacts on gastrointestinal health, gynaecological or reproductive health, short term and chronic pain and impacts on weight, as a result of child sexual abuse.[257]

---

[251] Risks include opportunities for offenders to access victim-survivors; to distribute CSAM; simulate abuse of virtual representations of children; and use integrated tech such as haptics, which simulate real world sensations such as movements, vibrations, and force. Source: WeProtect and the University of Manchester, 2023. Extended Reality technologies and child sexual exploitation and abuse. [accessed 12 June 2024].

[252] NSPCC (Allen, C.), 2023. Child Safeguarding and Immersive Technologies: An Outline of the Risks. [accessed 21 November 2024].

[253] NSPCC (Hamilton-Giachritsis, C., Hanson, E., Whittle, H. and Beech, A.), 2017. "Everyone deserves to be happy and safe". A mixed methods study exploring how online and offline child sexual abuse impact young people and how professionals respond to it. [accessed 11 June 2024].

[254] 5,862 victims and survivors participated in the Truth Project. Source: IICSA, 2022. The Report of the Independent Inquiry into Child Sexual Abuse. [accessed 22 September 2023].

[255] Maniglio, R., 2009. The Impact of Child sexual abuse on health: a systematic review of reviews. *Clinical Psychology Review*. 29(7), pp.647 – 657; Hailes, H.P., Yu, R., Danese, A., Fazel, S., 2019. Long-term outcomes of childhood sexual abuse: an umbrella review. *The Lancet Psychiatry.* 6(10), p.830 – 839.

[256] CSA Centre response to November 2023 Illegal Harms Consultation.

[257] CSA Centre response to November 2023 Illegal Harms Consultation.

2A. Grooming

Warning: This chapter contains content that may be upsetting or distressing in relation to grooming.

Summary of analysis for grooming offence: how harms manifests online and risk factors

This chapter explores the evidence on online grooming for the purpose of conducting child sexual abuse. Grooming is the process of building a relationship or emotional connection with a child or young person so they can manipulate, exploit, and abuse them.[258] Grooming is a complicated process that does not follow a set pattern. Sometimes it can take weeks or months, or it can happen quickly after the first contact. Grooming can involve many stages, and is not always limited to sexual conversations, although it can involve coercing or manipulating children into multiple sexual acts. The scale of online grooming is extensive, with the NSPCC reporting over 34,000 cases of online grooming crimes against children between 2017 and 2023.[259]Grooming will affect each victim and survivor differently, but the effects are serious and can often last a lifetime. These include negative psychological effects such as self-harm, loss of confidence, increased aggression and feelings of self-blame, and difficulties trusting others.

Service type risk factors:

Perpetrators can groom children in a many different ways, and through a variety of different services. **Social media, video sharing services, gaming services,** and **messaging services** are commonly used to carry out this offence. Other service types that children use, such as **discussion forums and chat rooms** are also involved in grooming cases.

Groomers can 'platform-hop' between services to exploit different functionalities, enabling them to meet children, engage in grooming conversations and move interactions with children to more private spaces to avoid detection.

User base risk factors:

For grooming offences, a high-risk factor is the **age of users.** Groomers seeking to target children are more likely to be drawn to services that children use. While grooming can occur on services of any size, those with a larger number or higher proportion of child users may offer more incentive for perpetrators during the early stages of grooming, such as identifying and making contact with children. This factor may become less important once contact is made. In the later stages of grooming, a service's perceived lack of detection technology may be a more

[258] HM Government,2021. Tackling Child Sexual Abuse Strategy 2021. [accessed 14 November 2024].
[259] IICSA, 2020. The Internet: Investigation Report. [accessed 31 August 2023].

important risk factor. However, a child's ability to access a service will remain a constant factor, given the nature of the offence.

Perpetrators assess a child's personal circumstances to identify traits that may make them more vulnerable to online grooming, such as low self-esteem or lack of supervision. Other factors that can increase vulnerability include identifying as **LGBTQIA +**, having a **disability**, **mental/physical health** and/or **socio-economic status**. **Gender** also appears to be a risk factor, with evidence showing that girls are significantly more likely to be groomed than boys, although male victims are believed to be under-reported. Importantly, none of these factors are the reason a child is sexually exploited. Rather, it is the way these circumstances can leave children more isolated or dependent on others, that makes them more susceptible to online grooming.

Functionalities and recommender systems risk factors:

Functionalities that allow abusers to identify and contact children are risk factors in enabling grooming offences. Perpetrators can use information on **user profiles** to find and target victims and survivors, initiating the grooming process. The ability to create **fake user profiles** allows them to misrepresent themselves by displaying a false age, name, or location. Studies also show perpetrators using **anonymous user profiles** to contact children.

User connections allow perpetrators to establish contact with children and start communication, while the sense of trust that mutual connections can create may also be exploited. Perpetrators may exploit network recommender systems, like publicly visible friends lists, to quickly infiltrate groups of children and use this access to blackmail and coerce them into further abuse. **User groups** can be used to target children, particularly groups for adolescents that discuss sexual themes.

Direct messaging enables perpetrators to make contact with children and potentially develop relationships through frequent communication, often away from public view. While **ephemeral and encrypted messaging** makes proving and detecting perpetrators' contact with children challenging.

Network recommender systems can facilitate grooming by suggesting adult users to children and vice versa. Additionally, users and groups associated with child users may be suggested to perpetrators, based on their past activity and connections.

Grooming can also often include coercing or manipulating a child into performing sexual acts over **livestreams**. Perpetrators can also use **comments on posted or livestreamed content** to build rapport and exchange contact details.

Other functionalities contribute to these offences. For example, **sending images through messaging functionalities** can be used to persuade children to share self-generated indecent imagery (SGII). **Visual media editing** functionalities can also be used to disguise the identity of perpetrators when they are contacting a child. The ability to **post or send location information** can allow a perpetrator details needed

to physically approach their target, to gain a child's trust by claiming common connections, or to use this knowledge to intimidate and threaten the child.

How grooming offences manifest online

2A.1 This section is an overview which looks at how the specified grooming offences manifest online, and how users may be at risk of harm.

Definition and Scale

2A.2 Grooming offences can include, but are not limited to, sexual or indecent communications with children, engaging in sexual activity with a child, or in the presence of a child, or coercing or inciting the child into sexual activity. The aim of online grooming is to manipulate or coerce children into engaging in sexual activity. This may occur solely online, through coercing children into producing and sharing self-generated indecent imagery (CSAM), or perpetrators may persuade children to meet them in person to sexually abuse the child.[260]

2A.3 The grooming process necessarily involves at least two components: identifying a child and contacting a child. Therefore, the presence of children on a service is a necessary initial enabler of grooming. While identifying a child may not constitute an offence, it is a crucial step towards committing grooming offences. The evidence presented here addresses both stages.

2A.4 The scale of online grooming is widespread, with IICSA describing it to be "*of real and significant concern*".[261] In 2023, the National Society for the Prevention of Cruelty to Children (NSPCC) reported nearly 34,000 recorded online grooming crimes against children over the past six years.[262] A US study of undergraduates found that 17% (nearly 1 in 5) had experienced sexual solicitation as youths from adults they had chatted with online, and 23% (nearly 1 in 4) recalled a long intimate conversation with an adult stranger, potentially indicating online grooming.[263]

2A.5 Unsolicited sexual messages that children receive from unknown users is also an indicator of online grooming. Ofcom research found that 13% (more than 1 in 10) of 11 to 18-year-olds had received pictures or videos of naked or half-dressed people when communicating online, and 10% had been asked to share an intimate picture or video of themselves.[264] A European study found that 68% (more than 3 in 5) of the 18-year-olds surveyed had experienced at least one sexual harm online during childhood, with 55% (more than half) being asked to engage online in sexually explicit activities they were uncomfortable with, or did not want to do.[265]

---

[260] Child Exploitation and Online Protection Centre (CEOP), 2022. What is sexual grooming?. [accessed 14 November2024].
s

[261] IICSA, 2020. The Internet: Investigation Report. [accessed 31 August 2023].

[262] Analysis of Freedom of Information requests sent to UK police forces: NSPCC, 2024. 82% rise in online grooming crimes against children in the last 5 years. [accessed 11 June 2024].

[263] The study was of 1,133 undergraduate college students at two public institutions in the United States and asked about their experiences when under 18. Greene-Colozzi, E., Winters, G., Blasko, B. and Jeglic, E., 2020. Experiences and Perceptions of Online Sexual Solicitation and Grooming of Minors. A Retrospective Report. *Journal of Sexual Abuse*, 29:7, 836-854.

[264] Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research. [accessed 14 November 2024].

[265] The study also found that 56% had received sexually explicit content from an adult they know or did not know: WeProtect, 2023. Estimates of childhood exposure to online sexual harms and their risk factors. [accessed 12 June 2024].

2A.6 There is evidence that incidences of grooming online are increasing. The NSPCC reported that more than 7,000 Sexual Communication with a Child offences were recorded by Police in 2023/24, an 89% increase since this offence came into force in 2017/18.[266]

2A.7 The scale of online grooming is likely underestimated due to complexities in how it presents, how it is experienced and challenges in identifying and reporting it, including the cross-border dimensions of offending where perpetrators and victims are often based in different countries. Such cases may be represented in these Police statistics, where perpetrators may not be UK citizens. Research has shown that grooming is significantly under-reported by victims and survivors for many reasons including shame, fear, and the lack of recognition that a crime has occurred. [267] Where it is reported, it is often many years after the abuse occurred.[268]

2A.8 Grooming is a behavioural offence that can take many different forms online and does not follow a set pathway. However, existing literature has identified some common patterns that occur in most cases. Children can be groomed to send CSAM and into being sexually abused online and/or offline through a single message, or over a prolonged period. Grooming can involve various stages, including forming friendships, flattery, developing trust, risk assessment, exclusivity, threats, and sexual conversations.[269] While these stages are not fixed, and can often occur rapidly and/or interchangeably, they indicate crucial points where interventions could potentially disrupt an offender's progression. The resulting abuse can be one-off, but can also often be sustained over a long period of time, with multiple offences being committed, and the abuse in some instances becoming increasingly severe.

2A.9 Grooming pathways often occur in stages which can take place on multiple different services.[270] Many perpetrators try to move children to more private online spaces to continue grooming, using functionalities such as private chat, end-to-end encryption, as well as image sharing, to carry out sexual communication.[271] Children are at risk at various points along the pathway, and different functionalities may present greater risks at different

---

[266] While these figures indicate an increase in grooming attempts, they reflect, at least in part, an increased national policing response to grooming, and so an increase in the detection and recording of offences that may already have been occurring in previous years. NSPCC, 2024. Online grooming crimes against children increase by 89% in six years. [accessed 14 November 2024].

[267] Quayle, E., Jonsson, L., Lööf, L., 2012. Online behaviour related to child sexual abuse. Interviews with affected young people. Council of the Baltic Sea States, Stockholm: ROBERT project. [accessed 31 August 2023]; Katz, C., Piller, S., Glucklich, T., & Matty, D. E., 2021. "Stop Waking the Dead": Internet Child Sexual Abuse and Perspectives on Its Disclosure. *Journal of Interpersonal Violence*, 36(9–10), NP5084–NP5104. [accessed 31 August 2023].

[268] Cited in Halvorsen, J. E. & Tvedt Solberg, E. & Hjelen Stige, S., 2020. "To say it out loud is to kill your own childhood." – An exploration of the first person perspective of barriers to disclosing child sexual abuse, *Children and Youth Services Review,* Elsevier, 113. [accessed 22 September 2023].

[269] Whittle, H. C., Hamilton-Giachritsis, E. and Beech, A. R., 2015. A comparison of victim and offender perspectives of grooming and sexual abuse, *Deviant behaviour*, 36 (7), pp.539-564; Borj P., Raja, K., & Bours, P., 2023. Online Grooming detection: A comprehensive survey of child exploitation in chat logs. *Journal of Knowledge Based Systems,* 259, 110039.

[270] The services used by perpetrators will differ according to the stage of the grooming journey. For example, the type of service where a perpetrator might seek to identify a potential child to groom may be a different from the one where they seek to exchange images with the child. Movement through the stages is often driven by the preparator's motives.

[271] Ofcom research found that 20% of 11-18-year-olds said they had communicated on more than one platform with the person with whom they had had their most recent potentially uncomfortable online contact experience. This 'uncomfortable experience' may not be grooming per se, but may be intimate image sharing, rude/abusive messages or being asked for personal information. The research also found that 10% of 11–18-year-olds said they had ever been asked to move their online conversation to another service by someone they did not know well or did not know at all: Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research; Ringenberg, T.R., Seigfried-Spellar, K.C., Rayz, J. M., and Rogers, M.K., 2022. A scoping review of child grooming strategies: Pre-and post-internet, *Child Abuse & Neglect*, 123, Article 105392.

stages. Thus, services must understand at what point in the grooming process their service could be exploited, to enable the implementation of appropriate safeguards.

2A.10 Some grooming perpetrators use online services to target large numbers of children, and one method deployed is the 'scatter gun' or 'pyramid approach'. This is where perpetrators engage with many children (sometimes hundreds), mostly unknown to them, in quick succession. In these instances, the desired outcome for the perpetrator is to obtain a response from a proportion of the targeted children to engage them in conversation, thereby beginning the grooming process.[272] This leads to a rapid escalation of the harm, with offences such as sexual communication sometimes being committed within minutes of the perpetrator and child making contact.[273]

2A.11 Other techniques used by perpetrators include relationship-building, such as 'the boyfriend model, where the course of the interaction can last for days or even years'.[274] Some perpetrators impersonate other young people or create fake online user profiles to build relationships with children and obscure their identity, using flattery or shared interests. In other cases, perpetrators may be truthful about who they are. Blackmail may also be used, which can make it difficult for the child to break contact with perpetrators of grooming or other CSEA offences online.[275]

Risks of harm to individuals presented by online grooming offences

2A.12 Grooming will affect each victim and survivor differently, but the effects are significant and long lasting.

2A.13 Grooming can lead to a range of negative psychological impacts. Qualitative studies based on interviews with victims and survivors of online grooming have reported impacts including self-harm, loss of confidence, aggression, and problems trusting others.[276]

2A.14 Feelings of self-blame are common as perpetrators often use manipulation and coercion to pressure victims into sending sexual material. This increases victims' and survivors' reluctance to disclose instances of online grooming.[277] The cycle of abuse can also be escalated, with perpetrators encouraging the child to include other children, objects, or

[272] Joleby, M., Lunde, C., Landström, Jonsson, L. S. 2021. Offender strategies for engaging children in online sexual activity, *Child Abuse & Neglect*, 120. [accessed 4 September 2023].
[273] Lorenzo-Dus, N., Izura, C., and Pérez-Tattam, R., 2016. Understanding grooming discourse in computer-mediated environments, *Discourse, Context & Media,* 12, pp.40-50. [*Note: this research involves the analysis of chat logs between perpetrators and adults posing as children. These may not be truly reflective of interactions between children and perpetrators.]*
[274] Barnardo's, 2017. Working with children who are victims or at risk of sexual exploitation: Barnardo's model of practice. [accessed 10 August 2023].
[275] Hanson, E, 2017. *The Impact of Online Sexual Abuse on Children and Young People: Impact, Protection and Prevention.* in (2017) *Online Risk to Children: Impact, protection and prevention* (First Edition ed.), Blackwell, John Wiley & Sons, pp.98-122.
[276] Whittle, H., Hamiliton-Giachritsis, C.& Beech, A., 2013. Victims' Voices: The Impact of Online Grooming and Sexual Abuse. *Universal Journal of Psychology* 1(2), pp.59-71. [accessed 18 November 2024]. Ofcom, 2024. Online communications among children and young people: Qualitative research exploring experiences of sexualised messages online.
[277] Hanson, E, 2017. *The Impact of Online Sexual Abuse on Children and Young People: Impact, Protection and Prevention.* in (2017) *Online Risk to Children: Impact, protection and prevention* (First Edition ed.), Blackwell, John Wiley & Sons, pp.98-122.

animals in the sexual abuse, deepening the child's complex feelings of shame and culpability.[278]

2A.15 Grooming can also affect victims' and survivors' attitudes towards being online. One study found that, rather than viewing the internet as a place of opportunity, victims' and survivors' attitudes shifted towards a more negative view.[279]

Evidence of risk factors on user-to-user services

2A.16 We consider that the risk factors below are likely to increase the risk of harm relating to grooming. These risk factors are summarised in the grey box at the start of the chapter.

Risk factors: Service types

2A.17 Although grooming offences can happen on various services, research shows that the following service types are commonly used to facilitate or commit grooming offences: **discussion forums and chat room services**, **social media and video-sharing services**, **private messaging services** and **gaming services**.

Discussion forums and chat room services

2A.18 There is a range of evidence that discussion forum and chat room services are used to identify and establish contact with a child, before conversation moves to a service with more privacy, such as a messaging service. The NSPCC noted that in reports of online grooming by children, conversations were said to start in a "*public online space such as a forum or group chat*" before "*becoming private*".[280] This is supported by evidence from Internet Matters, which showed that online groomers may strike up a relationship with a child through discussion forums before asking them to move to another service to talk privately.[281]

2A.19 Research suggests that young people who struggle to form friendships and relationships offline compensate by seeking online interactions, including in in chat rooms.[282] It is reasonable to assume that such an individual would be more vulnerable to approaches from an adult seeking to groom a child.

Social media services and video-sharing services

2A.20 Online grooming has been shown to take place on social media and video sharing services. In 2023, the NSPCC reported that, of the 34,000 online grooming crimes against children recorded in the last 6 years, where the means of communication was known, 26% took place on Snapchat and 47% took place on Meta-owned products such as Facebook and Instagram.[283] Protect Children also found that 48% of respondents who had sought contact with a child online first did so on social media; this was the most popular approach identified.[284]

---

[278] IICSA, 2020. The Internet: Investigation Report. [accessed 22 September 2023].
[279] Chiu, J. & Quayle, E., 2022. Understanding online grooming: An interpretative phenomenological analysis of adolescents' offline meetings with adult perpetrators. *Child Abuse & Neglect*, 128. [accessed 18 November 2024]
[280] NSPCC, 2020. The impact of the coronavirus pandemic on child welfare: online abuse. [accessed 10 August 2023].
[281] Internet.matters.org, n.d. Learn about online grooming. [accessed 10 August 2023].
[282] Wolak, J., Finkelhor, D., Mitchell, K. J., and Ybarra, M. L., 2008. Online 'Predators' and their victims, *American Psychologist*, 63(2) pp.111-128. [accessed 18 November 2024]
[283] NSPCC, 2024. 82% rise in online grooming crimes against children in the last 5 years. [accessed 11 June 2024].
[284] The survey was conducted among individuals searching for CSAM on dark web search engines: Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].

2A.21 Perpetrators may intentionally use social media and video sharing services to find and communicate with children, as a high proportion of children often use these services, an Ofcom report found that 60% of all children aged 3 to 17 game online.[285] WeProtect described how many perpetrators who attempt to groom children online first identify targets on social media before moving conversations to a more private space.[286]

2A.22 Perpetrators may also use social media services and video-sharing services due to functionalities that allow for quick and easy forming of many user connections (and subsequent communication), enabling techniques such as the 'scatter gun' or 'pyramid approach'. Some social media and video sharing services encourage new connections (see network recommender systems for more information), which perpetrators are known to exploit to quickly initiate contact with a child or children.

2A.23 Perpetrators may use social media services and video-sharing services to identify children, making use of live streaming functionalities (see live streaming for more information). Reports of online grooming made to the NSPCC mention children being approached on 'social media networks' and 'livestreaming platforms'.[287]

Messaging services

2A.24 There is evidence that messaging services are online spaces where perpetrators initiate conversations with children. Protect Children found that 37% of respondents who had sought contact with a child online first did so on a messaging app.[288] Data from the Office for National Statistics (ONS) has also shown that 74% of approaches to children online by someone they do not know first take place via messaging services.[289] In research from the NSPCC exploring the growth of online grooming reported during the COVID-19 pandemic, 'instant messaging apps' played a role – either as the first service used to contact a child, or in conjunction with other services used to communicate – in many of the reports where children discussed how the perpetrators had built relationships with them.[290]

2A.25 There is also evidence to show that perpetrators seek to move their communication with children to private messaging services, after initiating the conversation in a more public online space. A US study of online grooming by Thorn in 2022 found that 65% of children surveyed reported having an online-only contact invite them to move from a public chat into a private conversation on a different platform.[291] It noted that "*private messaging apps warrant unique consideration for the role they play in meeting people online and how these relationships deepen for minors*".[292]

---

[285] Ofcom, 2024. Children and Parents: Media Use and Attitudes Report. [accessed 24th September 2024].
[286] WeProtect, 2023. Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response. [accessed 12 June 2024].
[287] NSPCC, 2020. The impact of the coronavirus pandemic on child welfare: online abuse. [accessed 22 September 2023].
[288] The survey was conducted among individuals searching for CSAM on dark web search engines: Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].
[289] NSPCC response to November 2023 Illegal Harms Consultation.
[290] NSPCC, 2020.
[291] Half of the children surveyed (52%) reported having used a private messaging service to interact with an online-only contact, including 23% of the 9-12-year-olds, who had had daily interactions with an online-only contact using a private messaging service. 445 9–12-year-olds and 755 13- to 17-year-olds in the United States were surveyed. From this study it is not possible to infer that these were adult-led conversations, nor that grooming was involved: Thorn, 2022. Online Grooming: Examining risky encounters amid everyday digital socialization. [accessed 10 August 2023].
[292] Thorn, 2022. Online Grooming: Examining risky encounters amid everyday digital socialization. [accessed 10 August 2023].

2A.26 Perpetrators may seek to move their communication with children to messaging services to exploit the enhanced privacy of end-to-end encryption. WeProtect reported that many perpetrators first identify targets on social media, in chat rooms or in gaming environments before moving conversations to a private messaging app or an end-to-end encrypted environment to reduce the risk of detection.[293] For more information, see 'encrypted messaging' in the User communications sub-section.

Gaming services

2A.27 Evidence shows that gaming services are used by perpetrators to identify and establish contact with children. Protect Children found that 41% of respondents who had sought contact with a child online first did so via an online game.[294] NSPCC research on online grooming also highlights "*voice or text chat services built into online multiplayer games*" as methods used by perpetrators to approach children.[295]

2A.28 Gaming services may pose a particular risk because they are online spaces where communication with strangers is normalised, and often a core part of the gaming process, which may make children more vulnerable to approaches from perpetrators. Protect Children also noted that it is easier for perpetrators to hide their real identity in gaming, where many users do not reveal personal information.[296]

2A.29 CSEA professionals have also highlighted that gaming services can be exploited by those seeking to groom children, by using such services' features to gain contact and establish trust with them online.[297] Perpetrators have used in-game gifts and trades as a manipulation and coercion tactic, and instances of sexualised language and grooming have been observed in multi-player games.[298]

Risk factors: User base

User base size

2A.30 Both large and small user bases can pose risks for online grooming. On a large service, a perpetrator may use a scatter-gun approach, randomly targeting many due to the wider pool of potential victims. Services with a smaller user base may enable perpetrators to identify a victim more easily with specific characteristics or vulnerabilities.

User base demographics

2A.31 The following section outlines the key evidence on user base demographics and associated risks of harm, including those related to protected characteristics. Services should consider how these demographic factors may intersect. It should be noted that the picture of the full nature and extent of the grooming threat is incomplete due to under-detection and under-reporting of this crime; children may not report due to a lack of awareness that they have

---

[293] WeProtect, 2023. Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response. [accessed 12 June 2024].
[294] The survey was conducted among individuals searching for CSAM on dark web search engines: Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].
[295] NSPCC, 2020. The impact of the coronavirus pandemic on child welfare: online abuse. [accessed 22 September 2023].
[296] Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].
[297] Interpol, 2020. Threats and Trends Child sexual exploitation and abuse: covid-19 impact. [accessed 25 August 2023].
[298] 5Rights, 2020. *Risky by Design*. [accessed 26 September 2023]; Hamilton-Giachritsis, C., Hanson, E., Helen, W., Alves-Costa, F., and Beech, A., 2020. Technology assisted child sexual abuse in the UK: Young people's views on the impact of online sexual abuse. *Children and Youth Services Review*, 119.; Stonehouse, R., 2019. Roblox: 'I thought he was playing an innocent game', The BBC, 30 May. [accessed 2 October 2023].

been groomed, or due to feelings of fear, shame, or considering that they will not be believed. [299]

2A.32 However, the available data suggests that user base characteristics including **age, gender, disability, sexual orientation and gender identity** and **media literacy** could lead to an increased risk of harm to individuals.

Age

2A.33 Age is an important risk factor in grooming offences. Perpetrators seeking to target children are drawn to services that children use.[300]

2A.34 There is evidence to suggest that children above the age of 13 may be at greater risk of being contacted by strangers than younger children, possibly due to the increasing ownership of phones and social media accounts around this age.[301] Internet Matters found that the proportion of children contacted by strangers online rises from 18% of 11 to 12-year-olds, to 25% of 13 to 14-year-olds.[302]

2A.35 Nonetheless, younger children are at risk of online grooming. In 2023, the NSPCC reported that out of the 34,000 online grooming crimes recorded over the six years prior, 5,500 involved primary-aged school children as targets.[303]

Gender

2A.36 Girls have been shown to be at greater risk of experiencing grooming online. In 2023, the NSPCC reported that 83% of recorded online grooming offences over the past six years, where gender was known, were against girls.[304] Ofcom research found that girls aged 16-18 were more likely than other groups to have encountered all ten of the potentially uncomfortable experiences asked about. [305]

2A.37 Boys are less likely to report sexual abuse, often due to the perceived social stigma surrounding this type of crime, which can impact the accuracy of gender-related data. Evidence also suggests the type of sexual abuse experienced may vary by gender. A number of sources have found that a large portion of reports involving SGII derived from financially motivated sexual extortion (FMSE) involve boys.[306] Analysis by the Canadian Centre for Child

---

[299] Katz, C., Piller, S., Glucklich, T., & Matty, D. E., 2021. "Stop Waking the Dead": Internet Child Sexual Abuse and Perspectives on Its Disclosure. *Journal of Interpersonal Violence*, 36(9–10), NP5084–NP5104. [accessed 31 August 2023].
[300] Kloess, J. A., Hamilton-Giachritsis, C. E. and Beech, A. R., 2019. Offence Processes of online sexual grooming and abuse of children via internet communication platforms, *Sexual Abuse*, 31(1), pp.73-96. [accessed 18 November 2024].
[301] An Ofcom report found that 95% of children aged 12-15 have smartphones, while 59% of 8–11-year-olds do: Ofcom, 2024. Ofcom children and parents media literacy. [accessed 25 September 2024].
[302] Internet Matters response to November 2023 Illegal Harms Consultation.
[303] NSPCC, 2024. 82% rise in online grooming crimes against children in the last 5 years. [accessed 11 June 2024].
[304] Whittle et al. (2013) found that girls may be twice as likely to be groomed: Whittle, H. C., Hamilton-Giachritsis, C., Beech, A., and Collings, G., 2013. A Review of young people's vulnerabilities to online grooming, *Aggression and Violent Behavior*, 18, pp.135-146; NSPCC, 2024. 82% rise in online grooming crimes against children in the last 5 years. [accessed 11 June 2024].
[305] 24% of girls aged 16-18 had been asked to share intimate images or videos, compared to the average of 10% for all boys and girls aged 11-18. The uncomfortable experiences we asked about in the survey were: an unwanted friend or follow request; asked to share naked or half-dressed pictures or videos; asked to share personal information; a friend request from someone pretending to be someone else; pictures or videos of naked or half-dressed people; abusive, nasty, or rude messages, voice note or comments, asked to video call/chat with someone you have not spoken to before; asked to move your chat to a different app or platform by someone you don't know well or don't know at all; added to a group chat which includes people you don't know well or don't know at all; added to a group video call which includes people you don't know. Source: Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research.
[306] FBI National Press Office, 2022. FBI and Partners Issue National Public Safety Alert on Financial Sextortion Schemes. [accessed 22 September 2023]; IWF, 2023. The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 22 September 2023].

Protection (C3P) into FMSE found that, in cases where the victim's gender was known, 98% of financial sextortion victims and survivors were male.[307]

Disability

2A.38 Neurodivergent children, and those with disabilities such as learning difficulties, could be more vulnerable to online grooming. Ofcom research also found that those aged 11 to 18 with limiting or affecting conditions were more likely than those without such conditions to report potentially uncomfortable experiences asked about in the research, including being asked to share intimate images or being sent intimate images.[308] WeProtect also found that children with disabilities were more likely to experience online sexual harms in their childhood.[309] Whilst not specific to online harm, the CSA Centre found that disabled participants were twice as likely as non-disabled participants to describe experiences of CSA.[310] Professionals within charities working with children and young people who have experienced grooming and sexual abuse also highlight that children with special educational needs and disabilities (SEND) are considered to be at higher risk of engaging and responding to sexualised messages from adults.[311]

Sexual orientation and gender identity

2A.39 Evidence suggests that LGBTQIA+ children are at greater risk of online grooming, as they are potentially more likely to seek relationships online if they feel that they have little opportunity to explore their sexual orientation or gender identity offline. WeProtect found that respondents who identified as transgender/non-binary (59%) or LGBQ+ (65%) were more likely to experience online sexual harms in their childhood than those who identified as cisgender (47%) or non-LGBQ+ (46%).[312] Ofcom research found that those aged 11 to 18 who identified as LGBTQ+ were more likely than those who identified as heterosexual or cisgendered to encounter potentially uncomfortable experiences online.[313] Furthermore, research from US-based charity, Thorn, found that 32% (nearly one-third) of LGBTQ+ participants reported an online sexual interaction with someone they believed to be over 18, compared with 22% of non-LGBTQ+ participants.[314]

2A.40 There is also evidence to suggest LGBTQ+ victims and survivors face distinct barriers to disclosing and reporting child sexual abuse, leading to potential under-representation in reported cases. The IICSA found that reasons for under-reporting of abuse by LGBTQIA+ victims and survivors included internalised prejudice and stigma, and for gay and bisexual men, the continuing trauma of past criminalisation of homosexuality. [315]

---

[307] C3P, 2022. An analysis of financial sextortion victim posts published on R/Sextortion. [accessed 2 August 2023].
[308] Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research. [accessed 18 November 2024].
[309] WeProtect, 2021. Global Threat Assessment 2021. [accessed 22 September 2023].
[310] CSA Centre (Karsna, K. and Kelly, L.), 2021, The scale and nature of child sexual abuse: Review of evidence. [accessed 20 June 2024].
[311] Ofcom, 2024. Online communications among children and young people: Qualitative research exploring experiences of sexualised messages online
[312] WeProtect, 2021. Global Threat Assessment 2021. [accessed 22 September 2023].
[313] This included being asked to share intimate images or being sent intimate images: Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research. [accessed 18 November 2024].
[314] Thorn and Benenson Strategy Group, 2023. Youth Perspectives on Online Safety, 2022: an Annual Report of Youth Attitudes and Experiences. [accessed 12 June 2024].
[315] This report does not specify the online aspect of child sexual abuse, or grooming in particular, but given the prevalence of an online aspect in CSEA, it could be indicative of trends within online abuse: Independent Inquiry Child Sexual Abuse (Gibson, E., Knight, R. Durham, A. and Choudhury, I.), 2022. Engagement with lesbian, gay, bisexual, transgender and queer/ questioning + victims and survivors. [accessed 20 June 2024].

Socio-economic factors

2A.41 Perpetrators may seek out children who display certain vulnerabilities, such as being in care.[316] Whilst not specific to online harm, the CSA Centre found those who had lived in a care home were nearly four times as likely to have experienced CSA than those who had not.[317]

Physical/mental health

2A.42 Children with mental health difficulties may be at greater risk of online grooming. Evidence suggests perpetrators may seek out children who display certain vulnerabilities online, including mental health difficulties.[318]

2A.43 Children with experience of an eating disorder may be more vulnerable to online grooming, as there is evidence to suggest that some perpetrators deliberately target them. An exploratory study on children's vulnerability to human trafficking reported on a number of criminal and investigative journalism cases (across the UK, the Netherlands and Germany) where CSEA perpetrators posed as 'anorexia coaches'[319] to exploit sexual images and acts from young women and girls.[320] Children with experience of an eating disorder are more likely to connect with 'anorexia coaches' and are therefore at greater risk of harm of sexual abuse. For more information, see the 'Eating disorder content' chapter in Volume 3 of our May 2024 Protection of Children Consultation.

Ethnicity

2A.44 There may be underreporting of CSEA among children from minority ethnic groups, which may affect the accuracy of the ethnicity-related data for this harm. Police and local authority data indicate that children from these groups are not routinely identified as victims of sexual exploitation, suggesting that their risks and experiences of harm may be overlooked.[321]

2A.45 The pattern of under-reporting may be due to barriers such as cultural stereotypes, racism, shame, and stigma, which may place these victims and survivors at greater risk of not getting adequate support. The IICSA found that cultural stereotypes and racism had contributed to child sexual abuse going unrecognised or ignored by professionals. Additionally, societal racism can make individuals in ethnic minority communities hesitant to report abuse, fearing reinforcement of negative stereotypes. Furthermore, participants described how shame, stigma, and a fear of ostracisation created further barriers to speaking out. [322]

---

[316] Whittle, H. C., Hamilton-Giachritsis, C., Beech, A., and Collings, G., 2013. A review of young people's vulnerabilities to online grooming, *Aggression and Violent Behavior*, 18, pp.135-146; CSA Centre (Karsna, K. and Bromley, P.), 2023. Child sexual abuse in 2021/22: Trends in Official data. [accessed 31 July 2023].

[317] CSA Centre (Karsna, K. and Kelly, L.), 2021, The scale and nature of child sexual abuse: Review of evidence. [accessed 20 June 2024].

[318] Whittle, H. C., Hamilton-Giachritsis, C., Beech, A., and Collings, G., 2013. A review of young people's vulnerabilities to online grooming, *Aggression and Violent Behavior*, 18, pp.135-146.; CSA Centre (Karsna, K. and Bromley, P.), 2023. Child sexual abuse in 2021/22: Trends in Official data. [accessed 31 July 2023].

[319] These users are known as 'anorexia coaches', 'ana buddies' or a similar equivalent. 'Coaching' can include requesting pictures and videos for 'body checks', weekly weigh-ins and enforcing strict rules about what food to eat and avoid. It can also include 'punishment' for not complying in the form of verbal abuse, and sometimes requesting sexual images.

[320] Dettermeijer-Vermeulen, C., Esser, L. and Noteboom, F. 2016. Vulnerability up Close: An exploratory study into the vulnerability of children to human trafficking. [accessed 18 December 2023].

[321] Missing People, 2023. The ethnicity of missing people: Findings from police and local authority data, 2021-22. [accessed 17 June 2023].

[322] Many victims and survivors described how abuse had a significant effect on their sense of identity and belonging in the community they grew up, with some being ostracised from their communities and cut off from their friends and family. This report does not specify the online aspect of child sexual abuse, or grooming in particular, but given the prevalence of

Media literacy

2A.46 There is evidence to suggest that child users sometimes do not understand the risks associated with using a service, such as the risk of sharing personal information, and may not fully understand security settings.[323] In these cases, the lack of information from a service on how certain functionalities work may increase the risk of online grooming, leading to uninformed decisions (for example, child users may be more inclined to post personal information without understanding the personal risks that this may entail).

Risk factors: Functionalities and recommender systems

User identification

User profiles

2A.47 User profiles, and the information that is presented on them, facilitate the commission of grooming offences as they help perpetrators identify children to target. Perpetrators have described selecting potential victims based on information provided in user profiles, such as profile pictures, name, age, and location.[324] Malesky examined the online activity of 31 convicted sex offenders who had communicated with a child online and found that these offenders first viewed user profiles to identify potential victims.[325]

2A.48 User profiles can make information regarding the increased vulnerability of a child more visible to perpetrators and inform a perpetrator's risk assessment. NCMEC uses a model of online grooming which highlights that the first step many perpetrators take is a type of evaluation in which they seek to understand a child's personal characteristics.[326] This can include assessing whether the child's user profile indicates low self-esteem or a lack of supervision.

2A.49 Perpetrators can use multiple user profiles to re-victimise victims and survivors, and find new targets, by creating new accounts when their original account has been blocked, deleted, or removed. In response to Ofcom's Illegal Harms consultation, GeoComply described how perpetrators have been able to re-victimise children and find new children on platforms even when they have already been banned or removed by a platform.[327]

2A.50 The risk of user profiles intersects with fake user profiles, as perpetrators can have multiple accounts under fake profiles at any one time, irrespective of whether an account has been banned or removed.

---

an online aspect in CSEA, it could be indicative of trends within online abuse: Independent Inquiry Child Sexual Abuse (Rodger, H., Hurcombe, R., Redmond, T. and George, R.), 2020. "People don't talk about it": Child sexual abuse in ethnic minority communities. [accessed 20 June 2024].

[323] Wolak, J., Finkelhor, D., Mitchell, K. J., and Ybarra, M. L. ,2008. Online 'Predators' and their victims, *American Psychologist*, 63(2) pp.111-128. [accessed 18 November 2024].

[324] Quayle, E., Allegro, S., Hutton, L., Sheath, M. and Lööf, L., 2014. Rapid skill acquisition and online sexual grooming of children, *Computers in Human Behaviour*, 39, pp.368-375. [accessed 18 November 2024].

[325] The study found that offenders based their decision on who to contact on the presence of sexual content in a child's profile; an explicit statement of age; the perceived neediness or submissiveness of the child; and young-sounding usernames: Malesky, L. A., 2007. Predatory online behavior: Modus operandi of convicted sex offenders in identifying potential victims and contacting minors over the internet. *Journal of Child Sexual Abuse*, 16, pp.23–32. [accessed 18 November 2024]

[326] NCMEC (Australia) adapted the Winters & Jeglic Sexual Grooming Model for online perpetrators. Source: NCMEC, 2022. The new "Stranger Danger": Tactics used in the online grooming of children. [accessed 11 August 2023].

[327] GeoComply Solutions response to November 2023 Illegal Harms Consultation.

Fake user profiles and anonymous user profiles

2A.51 Evidence suggests the ability to create a fake user profile, which allows users to present false representations of themselves, is widely regarded as a key tool used by perpetrators to facilitate the commission of grooming offences. Abusers can create fake user profiles and present themselves as desirable to their target by falsifying their age, name and location.[328] This was recently seen in a high-profile case: a man was jailed in 2021 after posing as a teenage girl online and grooming 500 boys, blackmailing over 51 boys into sending indecent images of themselves, and coercing them into committing abusive acts against themselves and other children.[329] Ofcom research also found that 15% of 11 to18-year-olds claim to have received a friend request from someone pretending to be someone else.[330] Children and young people who have received sexualised messages from users perceived to be adults reported that accounts with heavily anonymised user profiles, for example with cartoon profile pictures and very few followers often used to send explicit images to them.[331]

2A.52 Perpetrators often use services that do not require them to disclose much personal information. This can include services that require only an email address or a username, which can often be easily falsified. networking

User connections

2A.53 The ability for users to connect with each other facilitates grooming, as it allows perpetrators to establish contact with child users and begin communicating, often making it a necessary pre-cursor to direct engagement with a child. Online services that utilise such functionality are implicated in significant numbers of cases of online grooming.[332]

2A.54 A specific use of user connection functionality by perpetrators is the 'scatter-gun' or 'pyramid approach' where perpetrators use user connection functionalities to try to access large numbers of children. One study provided an example of an offender randomly adding children to initiate contact with them.[333] Similarly, Internet Matters reported that social media services will often be used by perpetrators to target a large number of young users by sending out multiple connection requests.[334] Specific functionality such as 'quick add' has also been noted as reducing barriers to adults connecting with children.[335]

2A.55 The visibility of user connections may facilitate grooming as these features instil a sense of 'relationship' among mutual connections, which can be exploited by abusers to appear as a trusted contact of mutual friends. Young people's social networking predominantly involves

---

[328] Bluett-Boyd, N., Fileborn, B., Quadara, A. and Moore, S., 2013. The role of emerging communication technologies in experiences of sexual violence: a new legal frontier? [accessed 11 August 2023].
[329] BBC News, 2021. David Wilson: Sex offender who posed as girls online jailed for 25 years. 10 February. [accessed 11 August 2023].
[330] Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research. [accessed 18 November 2024]
[331] Ofcom, 2024. Online communications among children and young people: Qualitative research exploring experiences of sexualised messages online
[332] For example, in a sample of 641 grooming cases in England and Wales from 2020 where the online service was known, Instagram was used 236 times (37%), Facebook, Instagram and WhatsApp combined were used 324 times (51%), and Snapchat was used in 20% of cases: NSPCC, 2020. Instagram most recorded platform used in child grooming crimes during lockdown. [accessed 18 October 2024]
[333] Kloess, J. A., Hamilton-Giachritsis, C. E. and Beech, A. R., 2019. Offence Processes of online sexual grooming and abuse of children via internet communication platforms, *Sexual Abuse*, 31(1), pp.73-96. [accessed 18 November 2024].
[334] Internet.matters.org, n.d. Learn about online grooming. [accessed 10 August 2023].
[335] Young people interviewed about their experiences of receiving sexualised messages from adults (or perceived adults) felt this function enables adults to quickly add many users including children, widening the base of who they are engaging with. Source: Ofcom, 2024. Online communications among children and young people: Qualitative research exploring experiences of sexualised messages online

pre-established networks, and trust in the other users within the network. This sense of trust expands to those connected or 'friended' with others in the network [336] and can provide a false sense of security, for example, an abuser seeming to be 'known' to their social network.[337] Differences in the nature of sexualised messages received from users with mutual connections compared to those with none have also been highlighted. Children and young people discussing their experiences of receiving sexualised messages from users perceived to be adults suggested that messages from users with no mutual connections were often more likely to be sexualised or explicit in the first instance. Messages from users with some mutual connections were more likely to start off 'normal', before becoming sexualised over time.[338]

2A.56 As well as being used to identify targets, user connections can be used to coerce children. Perpetrators may use user connection lists to demonstrate to children that they know who their family and friends are, which can enable them to manipulate, threaten and coerce children, for example by threatening to share intimate images of the child with their contacts.[339]

User search

2A.57 Search functions that enable children, or groups containing children, to be found on services increase the risk of grooming. These functions allow perpetrators to identify potential victims based on specific characteristics. Research suggests that the ability to target particular groups of potential victims aids in the grooming process.[340]

User groups

2A.58 Online communities may facilitate grooming as they provide a space for an abuser to approach a child in a discreet way, whether as part of a user group discussing a shared interest, or as part of a sexualised environment among adolescents.[341] Evidence suggests there are online communities that are popular among adolescents that focus on explicit sexual discussions and obscene language [342] and engaging with other users in such an environment may desensitise users to sexual solicitations from perpetrators in these communities.

2A.59 User groups can be used by perpetrators to attract children. They can create groups centred around topics that appeal to young people to capture their interest, and may impersonate young people, potentially for several months to build trust.[343] Some perpetrators work together in such groups to enable grooming.

---

[336] Bluett-Boyd, N., Fileborn, B., Quadara, A. and Moore, S. 2013. The role of emerging communication technologies in experiences of sexual violence: a new legal frontier? [accessed 11 August 2023].
[337] Hamilton-Giachritsis, C., Hanson, E., Helen, W., Alves-Costa, F., and Beech, A., 2020.
[338] Ofcom, 2024.
[339] Safer Schools, 2023. Protecting Young People from Sextortion. [Accessed 1 September 2023].
[340] Bluett-Boyd, N., Fileborn, B., Quadara, A. and Moore, S., 2013. The role of emerging communication technologies in experiences of sexual violence: a new legal frontier? [accessed 11 August 2023].
[341] de Santisteban, P., Del Hoyo, J., Alcázar-Córcoles. M. Á.,, Gámez-Guadix, M., 2018. Progression, maintenance, and feedback of online child sexual grooming: A qualitative analysis of online predators. *Child Abuse Neglect*. 80, pp.203-215.
[342] Wolak, J., Finkelhor, D., and Mitchell, K. J., Ybarra, M. L., 2008. Online 'Predators' and their victims, *American Psychologist*, 63(2), pp.111-128.
[343] de Santisteban, P., Del Hoyo, J., Alcázar-Córcoles. M. Á.,, Gámez-Guadix, M., 2018. Progression, maintenance, and feedback of online child sexual grooming: A qualitative analysis of online predators. *Child Abuse Neglect*. 80, pp.203-215

User communications

Livestreaming

2A.60 Livestreaming allows users to share content in real time. This can be broadcast to limited connections, open to all users of a site or it can be a one-to-one live video.

2A.61 Livestreaming can be used by perpetrators to enable sexualised conversations with children and/or incite children to engage in sexual activity, sometimes including other children, in real time. The NSPCC found that sexualised conversations take place when children livestream and, of those children who livestreamed, 6% (more than 1 in 20) had received requests to change or remove their clothes. The research found that primary-school-aged children were more likely than secondary-school-aged children to be asked to change or remove their clothes when livestreaming.[344]

2A.62 Livestreaming functionalities in combination with screen-recording can enable perpetrators to create permanent records of SGII. Once a perpetrator has explicit photos or videos of a child, these can then be used to blackmail children into sharing further images[345]or to financially extort them.[346]

Video calling

2A.63 Video calling can be used by perpetrators to ask children to engage in sexual activity or to create sexually explicit material while on the call. An NSPCC study found that of the 40,000 children aged 7 to 16 who participated in the study, 12% (more than 1 in 10) of children have video-called with someone they did not know. During these calls, 10% of primary-aged and 11% of secondary-aged children were asked to change or remove their clothes.[347]

Direct messaging

2A.64 Direct messaging can facilitate online grooming by allowing private, direct, and rapid communication between perpetrators and children. This can allow perpetrators to build relationships away from public view and parental supervision.[348]

2A.65 The privacy of direct messaging can reduce barriers such as social status and age that are present in face-to-face environments. This increased privacy can increase feelings of intimacy and allows for more freedom to broach sensitive topics such as sex.[349]

2A.66 Evidence has also shown that children and young people are experiencing potentially uncomfortable experiences on direct messages. Ofcom research found that among those aged 11 to 18 who reported such experiences, messages—both individual and group—were the primary method of communication with the other user(s).[350] In qualitative research exploring the experiences of children and young people who had received sexualised

[344] Survey of nearly 40,000 children aged 7 – 16 years old: NSPCC, 2018. *Livestreaming and video-chatting.* [accessed 22 September 2023].
[345] NCA, Child sexual abuse and exploitation - National Crime Agency [accessed 23 September 2024].
[346] NSPCC 2024. Young people's experiences of online sexual extortion or 'sextortion'. [accessed 25 September 2024].
[347] NSPCC, 2018. Livestreaming and video-chatting. [accessed 22 September 2023].
[348] See case studies analysed in source: Kloess, J. A., Hamilton-Giachritsis, C. E. and Beech, A. R., 2019. Offence Processes of online sexual grooming and abuse of children via internet communication platforms, *Sexual Abuse*, 31(1), pp.73-96.
[349] Wolak, J., Finkelhor, D., and Mitchell, K. J., Ybarra, M. L., 2008. Online 'Predators' and their victims, *American Psychologist*, 63(2), pp.111-128.
[350] Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research.

messages from users perceived to be adults, all respondents had been contacted via direct messaging functionalities on various online services.[351]

2A.67 The ability to share images via direct messaging can enable grooming offences by encouraging children to share SGII. Evidence has shown that to normalise their requests for sexual images and to lower a child's inhibitions, perpetrators may first share pornographic images with the child.[352] This is followed by asking the child to send the perpetrator similar images of themselves.

2A.68 The use of emojis in direct messages can help perpetrators to incite children to engage in sexual activity. A study exploring discussions on LiveMe found that emojis were used to send sexually suggestive messages. Their analysis of over 39 million chat messages found that emojis were used as a form of communication to request sexually inappropriate and suggestive acts, such as the removal of clothes.[353] The emojis used in these instances were clothing-related emojis, hand gestures and tongue emojis.

Ephemeral messaging

2A.69 The use of ephemeral messaging has also been noted in instances where children have received sexualised messages from users perceived to be adults. Young people reported feeling that services with this functionality were more conducive to receiving sexualised messages form adults, and the self-deleting nature of the content meant that showing the messages to others, for example adults they trusted, was more difficult.[354]

Encrypted messaging

2A.70 As has been introduced previously, perpetrators may seek to move their communication with children to online spaces that offer end-to-end encryption, which can make it harder to detect offenders' contact with children. Law enforcement agencies have highlighted the impact that the increased prevalence of end-to-end encryption could have on detecting offenders and on child safety.[355] The National Crime Agency (NCA) also estimated that most reports (92% from Facebook and 85% from Instagram) that are currently disseminated to UK police each year will be lost as a result of the roll out of end-to-end encryption.[356]

2A.71 Perpetrators may be using end-to-end encrypted messaging to share tips and advice on harmful behaviour with each other. Protect Children reported that the lack of text moderation has allowed for offenders to share grooming strategies, discuss offending and share identified social media accounts of vulnerable children with each other.[357]

Group messaging

2A.72 Group messaging, like direct messaging, allows adults to engage directly with children and develop relationships. Ofcom research found that nearly 1 in 4 children (23%) had been

---

[351] Ofcom, 2024. Online communications among children and young people: Qualitative research exploring experiences of sexualised messages online

[352] Thomas, K., Hamilton-Giachritsis, C., Branigan, P. and Hanson, E., 2023. Offenders' approaches to overcoming victim resistance in technology-assisted child sexual abuse, *Child Abuse & Neglect,* 141(2).

[353] These chat messages were exchanged by more than 1.4 million users in 291,000 live broadcasts over two years: Lykousas, N. and Patsakis, C. 2021. Large-scale analysis of grooming in modern social networks. *Expert systems with applications, 176.*

[354] Ofcom, 2024.

[355] Virtual Global Taskforce, 2023. Statement on End-to-End Encryption. [accessed 16 August 2023].

[356] NCA, 2024. European police chiefs call for end-to-end encryption roll out to include public protection measures. [accessed 8 July 2024].

[357] Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].

added to a group chat with people they did not know well or at all, making it the second most common and potentially uncomfortable online contact experience reported.[358]

Commenting on content

2A.73 Perpetrators can use comments on posted content as a means of building rapport with victims and survivors in the early stages of the grooming journey. Reports by BBC News and The Times found the presence of sexually explicit comments on children's videos on TikTok and YouTube respectively.[359]

2A.74 There is also evidence to suggest that livestream comments are used to facilitate grooming offences, where hyperlinks or attempts to exchange contact details are shared in comments with the aim of getting the child to connect with the perpetrator on another service. As outlined in the Livestream section above, the NSPCC found that sexualised conversations take place when children livestream.[360] Transcripts from the IWF have also shown that perpetrators leave sexualised comments when children livestream.[361]

Posting or sending location information

2A.75 Sharing a user's current location can provide a perpetrator with the necessary information to physically approach their target. A service which automatically shares a child's location on shared content, or which gives a child the ability to post or send their location, can give perpetrators information about a child's frequent places, such as their school and home address. Perpetrators can use this information to gain the trust of children, or may use the knowledge to threaten the child, to further their abuse.

Transactions and offers

Accepting online payments

2A.76 Perpetrators may send money to a child or buy them gifts (either virtual or physical) to facilitate relationship building as part of the grooming process. This can be enabled by a service accepting online payments. Evidence shows that perpetrators give gifts to flatter children, and as a gesture to give the impression of affection to the child user.[362]

2A.77 Evidence also suggests that an increasing number of perpetrators are coercing children into sending SGII by offering them money. Thus, functionalities that allow children to accept online payments may facilitate this. Blackmailing child users into generating further sexual images was seen in the case of one perpetrator who posed as a rich businessman online and groomed children, inciting them to generate SGII with offers of financial payments. The perpetrator then threatened to distribute these images to the child's friends and family unless the child sent further indecent images.[363]

---

[358] Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research.
[359] Shukman, H. and Bridge, M., 2018. Paedophiles grooming children live on YouTube, *The Times*, 10 December. [accessed 21 September 2023]; Silva, M., 2019. Video app TikTok fails to remove online predators, *BBC News*, 5 April. [accessed 16 August 2023].
[360] NSPCC, 2018. Livestreaming and video-chatting [accessed 22 September 2023].
[361] King, J. 2022. The shocking transcripts that reveal how groomers sexually abuse children in their own rooms, *Metro,* 3 September 2022. [accessed 21 September 2023].
[362] Gámez-Guadix, M., De Santisteban, P., Wachs, S. and Wright, M., 2021. Unraveling cyber sexual abuse of minors: Psychometrics properties of the Multidimensional Online Grooming Questionnaire and prevalence by sex and age, *Child Abuse and Neglect*, 120.
[363] BBC News, 2021. Abdul Elahi: Sexual blackmailer jailed for 32 years.[accessed 16 August 2023].

Content storage and capture

Screen recording or capturing

2A.78 Screen-recording and screen-capture functionalities can be deployed by perpetrators during video calls or livestreaming to non-consensually capture indecent images.[364] They can then use these images to blackmail the child to generate further CSAM.

Content editing

Editing visual media

2A.79 Editing visual media can facilitate grooming offences by allowing perpetrators to disguise their identity, often posing as young people when first contacting children. They can do this by using video- or image-editing functionalities, such as filters, to change their appearance when calling a child, or when sending them photos of themselves. Deepfake technology and GenAI can assist in creating images and videos that may be uploaded to user-to-user services, making approaches to children appear more genuine and thereby increasing the likelihood of a response.[365] GenAI 'nudifying' apps can also be used to take a child's photos from their social media and create CSAM which can then be used to blackmail and 'sextort' them – for more information on these kinds of deepfakes, see the Intimate image abuse chapter, and for details on financial extortion resulting from media editing see 'financially motivated sexual extortion' in the introduction of this chapter.

Recommender systems

Network recommender systems

2A.80 Network recommender systems can play a role in facilitating grooming by suggesting child users to adults and adult users to children. Network recommender systems are used by services to recommend connections, and child users are likely to accept suggested connection recommendations, thereby expanding their online networks. Ofcom research found that 30% (nearly one-third) of children and young adults aged 11 to 18 said that they had added contacts via friend suggestions, quick adds, connections requests or cover functions.[366] Where an adult user has an established pattern of adding a lot of child contacts to their network, the recommender system could suggest this adult user to other children.

2A.81 A perpetrator may join various groups focused on topics that appeal to children, in an attempt to make contact with them. Recommender systems may then suggest further user groups or connections, based on the perpetrator's membership of these user groups.

---

[364] While users can often screen record or capture content using third-party services, screen recordings and captures are shared on U2U services as user-generated content and some U2U services have dedicated screen recording and screen capturing functionalities.

[365] Deepfakes are a specific type of media that involves the use of AI algorithms, particularly generative AI models, to modify videos, images or audio to create realistic synthetic content. This is often done by superimposing the face of a person onto the body of another person in a video or image as well as voice manipulation with lip syncing. Deepfakes are commonly shared as user generated content on user-to-user services but could also potentially be created using functionalities present on user-to-user services. Deepfake technology is currently used to create content that can be harmful; however, we acknowledge that it may also have positive use cases.

[366] Ofcom, 2023. Understanding Online Communications Among Children – Quantitative Research.

Risk factors: Business models and commercial profiles

Revenue models

2A.82 Services that generate revenue through the sale of online gifts or tokens may be at greater risk of being exploited by perpetrators of online grooming. The C3P have outlined how some individuals who perpetrate online grooming offences send gifts and tokens to children as part of the grooming process and/or to incentivise the child to share sexual imagery.[367]

2A.83 Platforms that facilitate money transfers or other U2U payments may be attractive for those seeking to perpetrate financially motivated sexual extortion schemes. Many services will generate revenue from each of these transactions.[368]

[367] Canadian Centre for Child Protection (C3P) response to November 2023 Illegal Harms Consultation.
[368] Canadian Centre for Child Protection (C3P) response to November 2023 Illegal Harms Consultation.

2B. Child Sexual Abuse Material (CSAM)

Warning: This chapter contains content that may be upsetting or distressing in relation to CSAM.

Summary analysis for CSAM: how harms manifest online, and risk factors

Online CSAM includes material depicting sexual activity, or indecent or prohibited imagery of children and can take the form of photographic images and videos, as well as non-photographic material, such as drawings and animations. CSAM can have a profound and long-lasting impact on children who are sexually abused, as well as on the wellbeing of adults and children who unintentionally view this material.

Beyond the abuse itself, the presence, sharing and viewing of images and videos depicting the abuse can serve as a continual source of trauma for victims and survivors of CSEA. Victims and survivors may experience re-victimisation and can be affected by heightened sensitivity to photos and cameras.

CSAM is often accessed intentionally, and the availability of such material online creates a permissive environment in which perpetrators may develop and act on their sexual interests. The availability of CSAM can lead to unintentional viewing, likely causing considerable distress.

Children themselves may generate content that can be considered CSAM, which can cause them harm. UK law enforcement refers to this as self-generated indecent imagery (SGII).

Service type risk factors:

Any service can be used to distribute CSAM. Services enabling image or video sharing, text posting or hyperlink sharing pose particular risks.

File-storage and file-sharing services, especially those that allow users to upload and share images through links, are considered particularly risky, as they enable large, curated collections of CSAM to be stored. While studies show that perpetrators also regularly encounter CSAM on **social media and video sharing services,**.

other types of services also pose risks of CSAM offences, including **discussion forums and chat rooms, messaging services** and **user-to-user pornography services**.

User base risk factors:

As outlined in the grooming section above, services with a **large user base** can pose a grooming risk and therefore may be considered risky for the creation of first generation CSAM. However, evidence suggests that perpetrators also often use

small and less-mature services to share CSAM, as these services may be less likely to have CSAM detection technology and processes in place.

Child users on a service can be a risk factor for CSAM, as offenders may search for content uploaded by children on their personal accounts, some of which may be considered CSAM. **Gender, disability, ethnicity, socio-economic factors** and **sexual orientation and gender identity** can factor in how likely children are to be vulnerable to the way the different CSEA offences manifest online.

Functionalities and recommender systems risk factors:

Several functionalities can facilitate CSAM offences. **Group messaging** enables CSAM sharing or trading within communities of users, while **direct messaging** and the ability to **post content,** such as text and images, are also used by perpetrators to share and distribute CSAM. **Encrypted messaging** enables perpetrators to share CSAM with less risk of discovery, while **ephemeral messaging** complicates detection due to the disappearing nature of the content. Messages or posts can include **hyperlinks** to collections of CSAM saved on file-storage and file-sharing services. These hyperlinks can be shared with perpetrators, sometimes for a fee. **Anonymous profiles** can allow perpetrators to avoid being personally identified by a service when sharing or accessing CSAM.

Livestreaming can allow abusers to create CSAM during livestream sessions or from SGII, which can then be widely distributed. This is particularly risky when combined with **storage** and **screen capture functionalities**.

The ability to **post goods or services for sale** can be exploited to distribute CSAM, while **cryptocurrency payments** pose a growing threat, as using cryptocurrencies allow offenders to buy CSAM anonymously and evade detection.

User profiles and **unauthenticated user profiles** can facilitate CSAM offences, by enabling abusers to target children. Additionally, they are also a tactic used to signpost other perpetrators to CSAM.

Functionalities allowing users to **download content**, such as CSAM, enable users to store and view local copies of content on their devices, as well as to share it with others.

Content **recommender systems** are also a risk factor in the viewing and discovery of CSAM, as a service's algorithm could suggest CSAM-related content to users who are actively viewing CSAM videos.

Business model risk factors:

Low-capacity services, and services that are **earlier in their business development** lifecycle will be at greater risk of being used by perpetrators to share CSAM. Early-stage services are less likely to have established processes or resources to detect and remove CSAM from their services. If a service has insufficient focus on having effective moderation and verification processes in place, this can be exploited by perpetrators to share CSAM content.

How child sexual abuse material offences manifest online

2B.1 This section provides an overview of how CSAM manifests online, and how users may be at risk of harm.

2B.2 Online CSAM includes material depicting penetrative sexual activity, non-penetrative sexual activity, or indecent or prohibited imagery of children.[369] This can take the form of photographic images and videos, as well as non-photographic material, such as drawings and animations. CSAM can also include deepfake imagery, 'pseudo-photographs'[370] and imagery created using generative AI tools accessed through extended reality technologies. CSAM is not limited to image-based material and can include materials that provide advice on grooming or abusing a child sexually, or material that is considered obscene and encourages the commission of other CSEA offences.[371]

2B.3 Perpetrators of CSAM may engage in creating, uploading, sharing, and distributing such material. They may also facilitate the creation, uploading, sharing, or distribution of CSAM by indicating that they possess it and by sharing links or advising others on where it can be found.

2B.4 While it is difficult to accurately estimate the volume of CSAM online, a number of sources show how widespread it is. NCMEC reported that it received over 36.2 million reports of suspected child sexual exploitation via its CyberTipline in 2023, a more than 12% increase on 2022, and the number of files included within the reports increased by 19% to more than 100 million.[372] The IWF confirmed that it received 275,652 reports containing CSAM, links to CSAM, or advertised CSAM in 2023.[373] Police data also shows that c.107,000 sexual offences against children were recorded by the police across England and Wales in 2022, a 7.6% increase on 2021 and a near quadrupling of the number recorded ten years prior. Police estimates suggest that online CSEA accounts for at least 32% of the recorded total.[374]

2B.5 Evidence suggests the presence of CSAM online is increasing. There have been year-on-year increases in the number of URLs which contain CSAM reported to the IWF, with an 8% increase between 2022 and 2023.[375]

2B.6 CSAM is present on both the dark and clear web. While CSAM can be found on the dark web, 97% of CSAM detected by the C3P Project Arachnid was hosted on the clear web.[376] Similarly, in a survey of individuals searching for CSAM on dark web search engines, 77% of

[369] Crown Prosecution Service, 2020. Indecent and Prohibited Images of Children. [accessed 18 August 2023].
[370] A pseudo-photograph is an image made by computer-graphics or otherwise which appears to be a photograph: Home Office, 2023. Indecent Images of Children: guidance for young people. [accessed 22 September 2023].
[371] For more information on what constitutes CSAM see the Illegal Content Judgement Guidance (ICJG - Volume 5, Chapter 26).
[372] NCMEC, 2024. CyberTipline 2023 Report. [accessed 19 November 2024]
[373] IWF, 2024. IWF Annual Report 2023 #behindthescenes. [accessed 8 July 2024].
[374] National Police Chiefs' Council (NPCC), 2024. National Analysis of Police-Recorded Child Sexual Abuse & Exploitation (CSAE) Crimes Report - January 2022 to December 2022. [accessed 20 September 2024]
[375] IWF, 2024. IWF Annual Report 2023 #behindthescenes. [accessed 8 July 2024].
[376] C3P, 2021. Project Arachnid: Online availability of child sexual abuse material. [accessed 21 August 2023].

respondents report that they have encountered CSAM or links to CSAM somewhere on the clear web.[377]

2B.7 The severity of harm is rising as more extreme categories of CSAM are detected, particularly involving babies and toddlers. In 2022, the IWF reported a yearly increase in Category A material, which includes images of penetrative sexual activity, sexual activity with animals, or sadism.[378] By the end of 2023, the IWF's hash database contained 10,393 unique image hashes of Category A material depicting children aged 0-2, and 554,553 Category A hashes in total, reflecting increases of 19% and 35%, respectively, compared to 2022. A hash serves as an indicator of the existence of such material.[379]

2B.8 Most of the CSAM presently detected on U2U services is content that has previously been shared, and sometimes reshared, hundreds of thousands of times over a period of many years.[380] In contrast, 'first-generation' or 'novel' CSAM refers to material that is newly generated and which has not been previously shared, re-shared or detected.[381] Perpetrators will often exchange first-generation CSAM in return for other new material.

2B.9 There is some evidence to suggest viewing abusive pornography can act as a gateway to perpetrators seeking out CSAM. The WeProtect Global Alliance's 2023 Global Threat Assessment reported on emerging evidence of an association between the frequent viewing of pornography and progression to viewing CSAM.[382] Similarly, the CSA Centre described how a common pathway into viewing CSAM involves initially viewing legal pornography, which gets more extreme and depicts younger individuals over time.[383] Interviews with offenders who have viewed CSAM in the UK also indicated that most had not intentionally sought out CSAM, but that it was a result of entrenched pornography use and spiralling online behaviour.[384] Further, the Lucy Faithfull Foundation has reported that of 3,400 callers to their Stop It Now Helpline in 2023, comprising adults and under 18s who said they had abused or were close to abusing, or were worried about their thoughts or behaviours, 25.9% (881) had self-reported a 'problem' with pornography.[385]

2B.10 Perpetrators actively identify and share legal content that is linked to CSAM using various tactics to indicate their ability to share it, but without publicly uploading it. This practice, known as 'contextual CSEA' includes sharing personal information about CSAM victims and

---

[377] Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].

[378] The IWF notes that of the reports it received in 2022, 255,588 were confirmed to have contained images or videos of children suffering sexual abuse. Of these, 51,369 were the most severe Category A images: IWF, 2023 The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 11 August 2023].

[379] IWF, 2023 The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 22 September 2023].

[380] This is in large part due to there being more consistent and developed systems for identifying known CSAM, such as the use of CSAM hashing databases, compared to identifying first-generation CSAM.

[381] This includes material that has been produced by a perpetrator who has sexually abused a child in person or who has directed the in-person sexual abuse of a child, or by a child creating 'self-generated' CSAM – known as 'self-generated indecent imagery' (SGII).

[382] We Protect Global Alliance, 2023. Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response. [accessed 10 June 2024].

[383] CSA Centre (Brown, S.), 2023. Key messages from research on child sexual abuse by adults in online contexts. [accessed 11 June 2024].

[384] The Police Foundation, 2022. Turning the Tide Against Online Child Sexual Abuse. [accessed 10 June 2024].

[385] Lucy Faithfull Foundation response to November 2023 Consultation, p.30

survivors to help locate material, as well as sharing contextual images taken from a sexual abuse 'series' to imply possession of illegal material.[386] Similar to tactics seen in other online harms, this practice—where perpetrators signpost others towards CSAM—is referred to as 'breadcrumbing'.

Risks of harm to individuals presented by child sexual abuse material offences

2B.11 For victims and survivors, knowing that CSAM remains available online and that perpetrators may still be using it, can be a continuing source of trauma. Some describe 'closure' as impossible.[387] Many describe feeling constantly in fear, and vulnerable, because their abuse exists as a permanent record online which others can view.[388]

2B.12 Survivors can be re-victimised by reliving their sexual abuse if they encounter the material online, or by the fear of being recognised by someone who has seen it. In a survey conducted by the C3P, 69% (more than two-thirds) of victims and survivors reported constant worry about being recognised, and almost a third had been identified online or in person by someone who had seen images of their abuse.[389] Some victims and survivors reported being targeted and re-victimised by someone who had recognised them, including being propositioned or threatened. Victims and survivors also describe suffering from a heightened sensitivity to photos and cameras.[390]

2B.13 The impacts of SGII, both non-consensual and aggravated,[391] can be wide ranging and severe. In the case of non-consensual SGII, the negative impacts on the child depicted are significant, often leading to mental health challenges[392] and negative social consequences,[393] particularly for girls, who will often face bullying, harassment, social exclusion, and victim-blaming.[394] Many victims and survivors of aggravated SGII describe feelings of self-blame, negative psychological health, and heightened anxiety from knowing that the images remain online.

2B.14 The re-victimisation of survivors can be exacerbated by the increasing use of generative AI to create CSAM. Children who have been abused may experience re-victimisation as offenders

---

[386] C3P, 2019. How we are failing children: Changing the Paradigm. [accessed 23 August 2023].
[387] CSA Centre (Brown, S.), 2023. Key messages from research on child sexual abuse by adults in online contexts. [accessed 11 June 2024].
[388] Owens, J. N., Eakin, J. D., Hoffer, T., Muirhead, Y., Lynn, J., & Shelton, E., 2016. Investigative aspects of crossover offending from a sample of FBI online child sexual exploitation cases. *Aggression and Violent Behaviour,* 30, pp.3–14.
[389] Sample consisted of a 150 victims and survivors. C3P, 2017. Survivors' Survey: executive summary 2017. [accessed 25 August 2023].
[390] NSPCC (Hamilton-Giachritsis, C., Hanson, E., Whittle, H. and Beech, A.), 2017. "Everyone deserves to be happy and safe". A mixed methods study exploring how online and offline child sexual abuse impact young people and how professionals respond to it. [accessed 31 August 2023].
[391] See the sub-section 'Cross-cutting harms' for a definition of non-consensual and aggravated SGII.
[392] Frankel, A., Bassm S., Patterson, F., Dai, T., Brown, D., (2018). Sexting, Risk Behaviour, and Mental Health in Adolescents: An Examination of 2015 Pennsylvania Youth Risk Behavior Survey Data. *Journal of School Health,* 88(3), pp.190-199.
[393] From a qualitative study with 41 young people in south-east England. Setty, E. 2019. A rights based approach to Youth Sexting: Challenging, Risk, Shame, and the Denial of Rights to Bodily and Sexual Expression Within Youth Digital Sexual Culture. *International Journal of Bullying Prevention,* 1, pp.298-311. [accessed 18 November 2024].
[394] Ringrose, J., Regehr, K., Whitehead, S., 2022. 'Wanna Trade?' Cisheteronormative homosocial masculinity and the normalisation of abuse in youth digital sexual image exchange. *Journal of Gender Studies*, 31(2), pp.243 – 261. [accessed 18 November 2024].

generate and distribute new sexual imagery of them.[395] Offenders also use generative AI to create new CSAM from innocuous images of children online, so many more children may be at risk of being newly victimised.[396]

2B.15 CSAM created using generative AI may also have other impacts, including the potential to normalise abuse and overwhelm both authorities and platforms. The NSPCC has warned that such use of generative AI risks normalising the sexual abuse of children.[397] The Australian eSafety Commissioner highlighted that as technology advances and enables the creation of increasingly realistic images, it may become harder for police forces and hot lines to identify children who are currently being abused and in need of urgent protection.[398] The large scale at which AI-generated CSAM can be produced presents a major and complex challenge to detection and moderation.[399] The NSPCC has also reported that some children have expressed nervousness at reporting AI generated images of themselves, or to speak to trusted adults, fearing they may not be believed when explaining that the images are artificially generated.[400]

2B.16 CSAM also has a broader impact on the population, as the unintentional viewing of CSAM – by both adults and children – is likely to cause considerable distress. 6% of British adults' report having been exposed to CSAM online.[401] A Childline report using data from UK counselling sessions found that young people who had accidentally accessed CSAM online often felt reluctant to confide in anyone about it. They feared they might not be believed or, that they could face arrest. The NSPCC found that "*some young people were so concerned about the repercussions of seeing this material that they were unable to sleep or were having anxiety attacks*".[402]

2B.17 Unintentional viewing of CSAM by adults may also risk creating a pathway to viewing more CSAM.[403] For some, it may cause individuals to become desensitised to the material and fall into more regular, intentional viewing that they then find difficult to stop.[404] A survey conducted by Protect Children, with people who view CSAM, found that 50% of respondents wanted to stop viewing CSAM.[405]

---

[395] Lucy Faithfull Foundation, 2024. A call to end AI-generated child sexual abuse. [accessed 11 June 2024].
[396] In a legal decision on deepfake CSAM in Canada, the court noted how this technology could be used to victimise any child using photos stolen from social media or taken surreptitiously in public: C3P response to November 2023 Illegal Harms Consultation.
[397] NSPCC response to November 2023 Illegal Harms Consultation.
[398] eSafety Commissioner, 2023. Generative AI – position statement. [accessed 12 June 2024].
[399] NSPCC response to November 2023 Illegal Harms Consultation.
[400] NSPCC response to May 2024 Protection of Children Consultation
[401] The figure is higher for young adults, with 14% of 18-24-year-olds reporting having been exposed to CSAM online: IWF, 2022. More than one in 10 British young people exposed to online child sexual abuse. [accessed 25 August 2023].
[402] NSPCC, 2016. Online child sexual abuse images: Doing more to tackle demand and supply. [accessed 25 August 2023].
[403] Over half (51%) of respondents to a survey of CSAM users on the dark web reported that they had first encountered CSAM accidentally, meaning they were exposed to CSAM without actively searching for it. Insoll, T., Ovaska, O. & Vaarenen-Valkonen. 2021. CSAM Users in the Dark Web: Protecting Children Through Prevention. [accessed 24 September 2024].
[404] Insoll, T., Ovaska, A. K., Nurmi, J., Aaltonen, M., and Vaaranen-Valkonen, N., 2022. Risk Factors for Child Sexual Abuse Material Users Contacting Children Online: Results of an anonymous multilingual survey on the dark web, *Journal of Online Trust and Safety*, 1(2). [*Note: This research was carried out on the dark web, which is out of scope of the Act*].
[405] Sample of 3,935: Suojellaan Lapsia, Protect Children (Insoll, T., Ovaska, A., and Vaaranen-Valkonen, N.), 2021. CSAM Users in the dark web: Protecting children through prevention. [accessed 25 August 2023].

Evidence of risk factors on user-to-user services

2B.18 We consider that the risk factors below are liable to increase the risks of harm relating to CSAM. These are also summarised at the beginning of the chapter.

2B.19 The evidence used in this chapter is not necessarily tied to individual offences, but the analysis relates more broadly to perpetrator actions that can lead to CSAM being created and appearing on U2U services. The nature of CSAM offences is such that the presence of CSAM online is likely to be very closely linked to the offences of creating, possessing, distributing, and publishing CSAM.

Risk factors: Service types

2B.20 While any service that allows users to share images, videos or text can allow perpetrators to distribute CSAM, the following types of services in particular can be used to facilitate or commit offences related to CSAM: **discussion forum and chat room services**, **social media services and video-sharing services, messaging services**, **file-storage and file-sharing services**, and **online user-to-user pornography services**.

Discussion forums and chat-room services

2B.21 Discussion forums can be used to embed and advertise CSAM. The IWF reported that 5% of the child sexual abuse imagery it detected appeared on forums.[406] In a separate 2018 study, specifically examining captures of livestreamed child sexual abuse, the IWF found that 73% (nearly three-quarters) of the images it discovered were embedded into 16 forums that were "*dedicated to the distribution of captures of live-streamed child sexual abuse*". These forums were "*at the centre of distribution networks for captures of live streamed child sexual abuse*".[407] [408]

Social media services and video sharing services

2B.22 There is a range of evidence that shows CSAM is available on social media and video sharing services. In a survey of individuals searching for CSAM, 29% (nearly one-third) of respondents said they had encountered CSAM on a social media platform.[409] Data from the NCMEC CyberTipline showed that, listed in order of level of reports, Facebook, Instagram, Twitter/X, Snapchat and TikTok accounted for 87% of all reports made in 2023.[410] Furthermore, the NSPCC reported that, among cases where a social media site was identified, Snapchat accounted for 48% of the offences relating to the sharing and

[406] IWF, 2023 The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 22 September 2023].
[407] Based on a sample of 2,082 images over three months: IWF, 2018. Trends in Online Child Sexual Exploitation: Examining the distribution of livestreamed child sexual abuse. [accessed 22 September 2023].
[408] Discussion forums and chat rooms may appear as features within a service. This could in some cases present a greater risk as these services may not have the user-to-user interactions that they enable as a primary focus and may lack robust moderation or trust and safety procedures.
[409] This was the second most popular location, behind pornography websites (32%): Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].
[410] NCMEC, 2023. . [accessed 23 September 2024].

possession of indecent images of children. Facebook and Messenger accounted for 10% of such offences, Whatsapp for 12% and Instagram for 6%.[411]

2B.23 There is further evidence that demonstrates perpetrators are actively using social media platforms to search for and share CSAM. Protect Children found that, in a survey of individuals searching for CSAM, 32% (nearly one-third) of respondents said they had used social media platforms to search for, view, or share CSAM.[412]

2B.24 Social media services can be also used to find information on how to commit child sexual abuse offences, and signpost users to other services where CSAM is shared. In a survey of individuals searching for CSAM, 16% of respondents said they learnt how to access CSAM in the dark web on social media platforms.[412]

Messaging services

2B.25 There is evidence to suggest CSAM is being encountered on messaging services.[413] Protect Children reported that, in a survey of individuals searching for CSAM, 12% had encountered CSAM on messaging apps.[414]

2B.26 There is further evidence to show that messaging services are being actively used by perpetrators to search for, view and share CSAM. One survey reported that messaging services are one of the online channels that perpetrators use to distribute CSAM.[415] Protect Children also reported that 29% of respondents had used a messaging app to search for, view, or share CSAM.[416]

2B.27 The use of messaging services may have increased since the pandemic. Research by Interpol showed that there was an increase in the volume of CSAM sent via private messaging applications, as well as social media services, during the COVID-19 pandemic.[417]

2B.28 Messaging can be used in conjunction with other services to facilitate CSAM offences; for example, to communicate with victims or facilitators of livestreamed CSEA.[418] Some offenders start this interaction via messaging, then move to another service to watch the livestream.[419]

---

[411] This is according to freedom of information data obtained from UK police forces, relating to 2021/22: NSPCC, 2023. We're calling for effective action in the Online Safety Bill as child abuse image crimes reach record levels. [accessed 11 June 2024].

[412] Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].

[413] Messaging services are defined in the glossary as: A user-to-user service type describing services that are typically centred around the sending and receiving of messages that can only be viewed or read by a specific recipient or group of people.

[414] Protect Children (Lapsia, S.), 2024.

[415] Lee, H. E., Ermakova, T., Ververis, V. and Fabian, B., 2020. Detecting child sexual abuse material: A comprehensive survey. *Forensic Science International*: *Digital Investigation*, 34. See 'Risks of harm to individuals presented by CSAM offences' for more information.

[416] The survey was conducted among individuals searching for CSAM on dark web search engines: Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].

[417] Interpol is an international organisation that facilitates police cooperation on international crime: Interpol, 2020. INTERPOL report highlights impact of COVID-19 on child sexual abuse. [accessed 20 June 2024].

[418] Napier, S., Teunissen, C. and Boxall, H., 2021. Live streaming of child sexual abuse: An analysis of offender chat logs, Trends and issues in crime and criminal justice, 639. [accessed 25 August 2023].

[419] Napier, S., Teunissen, C. and Boxall, H., 2021.

2B.29 Messaging services with end-to-end-encrypted messaging can make the exchange of CSAM harder to detect. For more information, see Risk factors: Functionalities and recommender systems and End-to-end encrypted messaging.

File-storage and file-sharing services

2B.30 File-storage and file-sharing services[420], particularly services that allow users to upload and share access to images and videos, present a significant risk for hosting CSAM because perpetrators can store it on these services. In 2023, INHOPE reported that approximately 39% of the CSAM it detected was hosted by 'image hosts'. They also reported that 5% of the CSAM detected was hosted by 'file hosts', although this figure has been as high as 26% in previous years. INHOPE suggested that this reduction from previous years was likely to be due to difficulties in detecting illegal content on file hosting services that require payment.[421] The IWF found that 89% of images or videos detected of livestreamed child sexual abuse were stored on an 'image-hosting service'. Other types of file-storage and file-sharing services such as 'cyberlockers' and 'image stores' made up 4% and 1% of cases respectively of the child sexual abuse imagery that the IWF reviewed in 2023.[422]

2B.31 File-storage and file-sharing services may also enable the sharing of CSAM, as perpetrators can distribute URLs directing users to these collections.[423] A study reported that this is done through URLs which are shared to services such as image boards, offender forums and 'chats'.[424] According to the IWF, "*image hosts allow users to upload still images which are assigned a unique URL and can be embedded to display on third-party websites, such as forums or social networking sites*".[425]

User-to-user pornography services and dating sites

2B.32 Perpetrators are encountering CSAM on user-to-user pornography services. Protect Children found that, in a survey of individuals searching for CSAM, roughly one third (32%) of perpetrators had encountered CSAM on pornography websites; making it the most common type of online service identified in the survey for where CSAM was encountered on the clear web.[426]

2B.33 Evidence suggests perpetrators are also using user-to-user pornography services to disseminate CSAM. A 2021 report by the C3P stated that Serverel, which is the hosting

---

[420] File-storage and file-sharing services are defined in the glossary as: 'User-to-user service type describing services whose primary functionalities involve enabling users to store digital content and share access to that content through links.'
[421] INHOPE are a global network of organisations working to tackle CSAM: INHOPE, 2023. Annual report 2023. [accessed 7 August 2024].
[422] IWF, 2023. Annual report 2023. [accessed 7 August 2024].
[423] We understand that perpetrators may be more likely to choose these services if they are encrypted, or if access is time limited.
[424] GCHQ Government Communications Headquarters, "A thematic overview of how the internet facilitates the distribution of Child Sexual Abuse Material." GCHQ Government Communications Headquarters, 2022. As cited in Dorotic. M. and Johnsen, J. W., 2023. Child Sexual Abuse on the Internet. Report on the analysis of technological factors that affect the creation and sharing of child sexual abuse material on the Internet. [accessed 25 August 2023].
[425] IWF, 2018. Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse. [accessed 22 September 2023]; IWF found that 77% of child sexual abuse imagery appeared on imaging hosts: IWF, 2023. The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 22 September 2023].
[426] Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].

provider for at least 1,200 unique websites sharing adult content, received 66,824 removal notices for hosting post-pubescent CSAM. Overall, more than 72,000 pieces of content were targeted for removal on Serverel sites (including both pre- and post-pubescent CSAM).[427]

2B.34 Perpetrators also use online user-to-user pornography services to seek out CSAM. This is evidenced by the chatbot released on Pornhub, which is used to intercept searches using known CSA search terms. The chatbot was launched in March 2022, and was used in 173,904 search attempts in the first 30 days after its launch.[428]

2B.35 Dating services can be exploited to solicit livestreamed abuse of children. A 2022 Australian study of 9,987 people who had used mobile or website dating platforms in the past five years, found that 12.4% of respondents reported receiving requests to exploit their own children or children they had access to.[429] These requests included seeking sexual information, images, or videos of children.

User base size

2B.36 Evidence suggests that both large and small services can pose a risk of CSAM. Some of the most prolific sharing of CSAM occurs in services with large user bases. In addition, evidence suggests that perpetrators seeking SGII will often target services with larger user bases. As described earlier in this chapter, perpetrators seeking to groom children for the purposes of creating SGII will often use services with a larger user base, allowing them to target a larger number of children using the 'scattergun approach'.[430]

2B.37 However, larger service sizes and user bases may correlate with greater detection and removal efforts, and as a result, the rates of detected CSAM on smaller services (with lower detection capabilities) may not be representative of the volume of CSAM present on those services. Intelligence suggests that perpetrators often seek out services with smaller user bases, particularly services that are less mature, as these services may have fewer CSAM detection technologies or processes in place. In addition, some services with a smaller user base offer users' specific functionalities which may not be available on services with larger user bases, such as the ability to post content without a registered account. Perpetrators may target these services to exploit such functionalities.

User base demographics

2B.38 The following section outlines the key evidence on user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex, and involve multiple factors.

2B.39 Data suggests that user base characteristics including **age**, **gender, ethnicity, socio-economic factors, disability, sexual orientation, and gender identity** could lead to an increased risk of harm to individuals.

---

[427] C3P, 2021. Project Arachnid: Online availability of child sexual abuse material. [accessed 21 October 2024].
[428] IWF, 2022. Internet Watch Foundation, Stop It Now, and Pornhub launch first of its kind chatbot to prevent child sexual abuse. [accessed 30 August 2023].
[429] Teunissen, C., Boxall, H., Napier, S. & Brown, R. 2022. The sexual exploitation of Australian children on dating apps and websites. [accessed 24 September 2024].
[430] NCA, 2021. National Strategic Assessment of Serious and Organised Crime. [accessed 31 August 2023].

Age

2B.40 Services with a young user base may be at increased risk of CSAM offences. Content uploaded by children may be sought out by potential perpetrators, and this content may be classifiable as CSAM. Further, as is discussed above, perpetrators can target services with a younger user base to identify children for the purposes of grooming, which may result in the production and sharing of CSAM, including SGII.

2B.41 Analysis by the IWF indicates that children of all ages are at risk. In a review of unique indecent image hash matches by age for 2023, the youngest child in the image was assessed to be aged 0-2 in 35,153 instances, with 227,437 images for children aged 3-6, and over 900,000 instances for both children aged 7-10 and those aged 11-13. There were also 96,322 unique indecent images of children where the youngest child was assessed as being aged 14-15, and 45,638 of children aged 16-17.[431]

Gender

2B.42 Gender is a risk factor for CSAM as most children depicted in CSAM are girls. The IWF found that 96% of the reports processed in 2022 depicted exclusively girls. Furthermore, many of the SGII reports (64%) received by the IWF in 2022 related to girls aged 11 to 13.[432]

2B.43 While most children depicted in CSAM are girls, there is some evidence to suggest that the content that does depict boys tends to be more severe. The IWF found that content depicting boys tended to be of higher severity (based on CSAM categories) than content depicting girls.[433]

2B.44 In terms of offender demographics, most perpetrators are male. The IWF found that, of the CSAM they analysed, where a perpetrator was visible, they tended to be male.[434] One study by the United States Sentencing Commission also found that 94.3% of offenders involved in CSAM production were male.[435]

Ethnicity

2B.45 Evidence has shown that children from a wide range of ethnic backgrounds are at risk. A study by ECPAT International and Interpol found that, of the analysed CSAM in their study in which ethnicity was determinable, both the majority of children (76.6%) and the majority of perpetrators (78.8%) were white.[436] In terms of other ethnicities, 10.1% of children were classified as Hispanic or Latino, 9.9% were Asian, and 2.1% were Black. For perpetrators, the research found that 12.2% were Hispanic or Latino, 4.2% were Black and 3.2% were Asian.

2B.46 As has already been described, there is evidence to suggest there may be an under-reporting of children from minority ethnic groups being identified as victims of CSEA, which may affect

---

[431] Unique Image Analysis | IWF 2023 Annual Report [accessed 20 November 2024]
[432] IWF, 2023 The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 22 September 2023].
[433] IWF, 2023. The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 22 September 2023].
[434] IWF, 2022. The Annual Report 2021. [accessed 30 August 2023].
[435] United States Sentencing Commission, 2021. Federal Sentencing of Child Pornography: Production Offenses. [accessed 22 September 2023].
[436] ECPAT and Interpol, 2018. Towards a global indicator on unidentified victims in child sexual exploitation material: summary report. [accessed 30 August 2023].

the accuracy of the ethnicity-related data for this harm. For more information, see Grooming, Risk factors: User base, ethnicity.

Socio-economic factors

2B.47 The socio-economic background of children, particularly when intersecting with gender, has been found to increase the risk of SGII occurring. Evidence has found that girls, particularly from less privileged backgrounds, are at greater risk of experiencing the non-consensual sharing of SGII.[437]

Disability

2B.48 Children with a disability may be more likely to be depicted in CSAM. Research indicates that perpetrators target and exploit the vulnerability of children with disabilities in order to sexually abuse them,[438] which can result in the production of CSAM. It is estimated that children with disabilities are nearly three times as likely to be sexually abused than children without disabilities.[439] In particular, research has found that children with disabilities or who are neurodivergent are more vulnerable to pressure from others to produce SGII.

Sexual orientation and gender identity

2B.49 LGBTQIA+ young people may be less open about and less supported with their feelings, and their isolation may make them more vulnerable to sexual exploitation, including pressure from others to produce SGII. [440]

2B.50 As has already been described, there is evidence to suggest there may be an under-reporting of LGBTQIA+ individuals as being identified as victims of CSEA, which may affect the accuracy of reported data for this demographic for this harm. [441] For more information, see 'Grooming, Risk factors: User base - sexual orientation and gender identity.'

Risk factors: Functionalities and recommender systems

User identification

User profiles

2B.51 User profiles and the accompanying statements on them, such as biographies, can facilitate CSAM offences. A study into TikTok, which is used predominantly by younger users, found that some TikTok user profiles included statements of interest in naked images and the exchange of sexual videos.[442]

---

[437] Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 30 August 2023].
[438] Independent Inquiry Child Sexual Abuse, 2022. Child sexual exploitation by organised networks: Investigation Report. [accessed 30 August 2023].
[439] Vera Institute of Justice (Smith, N. and Harrell, S.) 2013. Sexual Abuse of Children with Disabilities: A National Snapshot. [accessed 30 August 2023].
[440] Independent Inquiry Child Sexual Abuse, 2022. The Report of the Independent Inquiry into Child Sexual Abuse. [accessed 25 September 2024].
[441] Independent Inquiry Child Sexual Abuse, 2022.
[442] Cox, J, 2018. December 6. TikTok, the app super popular with kids, has a nudes problem. *VICE,* 6 December. [accessed 20 August 2023].

2B.52 The ability to create multiple user profiles can enable perpetrators to overcome measures such as strikes and blocking, by creating multiple profiles from which they access and share CSAM.

Fake user profiles

2B.53 Fake user profiles can be created by perpetrators to impersonate victims and survivors, so that perpetrators can connect with each other and share advice. The NSPCC has warned of 'tribute sites' and 'tribute user profiles', that impersonate victims and survivors so that abusers can connect with each other and share advice.[443] This is a CSEA breadcrumbing technique that facilitates the commission of CSAM offences.

Anonymous user profiles

2B.54 Anonymous user profiles, and the ability for unregistered users to post content anonymously, may allow perpetrators to avoid being personally identified by a service when sharing or accessing CSAM, and thereby avoid any potential content escalation or legal investigation. This may mean perpetrators are likely to choose services that allow users to share content without registering through creating an account, as this affords users a greater degree of anonymity. This was identified in a study into technologies used to commit child sexual abuse offences: 82% (more than 4 in 5) of the offenders surveyed indicated that anonymity was of at least moderate importance for conducting CSAM offences.[444]

2B.55 Perpetrators may leverage various forms of location-altering technology to hide their location and identity, to enable them to conduct illicit activities anonymously and therefore evade detection. GeoComply have highlighted how cybercriminals are using anonymising technology, such as Virtual Private Networks (VPNs), to commit sexual offences against children online.[445]

User networking

User groups

2B.56 Perpetrators can take advantage of user groups and use them as spaces to exchange CSAM, as well as ideas, advice, and tradecraft tips regarding abusive behaviour. The NSPCC noted that CSAM can be shared through online communities, and this behaviour can become "*normalised or even encouraged"* as like-minded people who share a sexual interest in children connect online.[446] For some closed user groups, sharing CSAM, including new, first-generation CSAM, is sometimes a condition of entry. Protect Children also found that perpetrators were joining thematic communities online that posted and traded violent material.[447]

[443] NSPCC response to Ofcom 2022 Call for evidence: First phase of online regulation.
[444] Steel, C., Newman, E., O'Rourke, S. and Quayle, E., 2022. Technical Behaviours of Child Sexual Exploitation Material Offenders, *Journal of Digital Forensics, Security and Law*, 17(1). [accessed 18 November 2024]
[445] VPNs are Virtual Private Networks: GeoComply Solutions 2024 response Ofcom Illegal Harms consultation.
[446] NSPCC, 2019. Online abuse: learning from case reviews. [accessed 30 August 2023].
[447] Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].

2B.57 In addition to sharing image CSAM directly within user groups, perpetrators also share links. Large virtual communities of offenders have been seen to share millions of items of CSAM indirectly via hyperlinks.[448]

2B.58 There is evidence to suggest user groups are forming that are dedicated to the abuse of AI tools to generate CSAM, which may increase the rate and ease by which the technology is exploited for these purposes. Newly created sections of online forums have been observed where members advise and request information on acquiring child sexual abuse-related material from AI systems.[449]

User communications

Livestreaming

2B.59 As has been introduced previously, livestreamed CSEA is a widespread problem, both in the UK and globally. This is where offenders view, comment on and direct the sexual abuse of children, in real time. It could be through one-to-one conversations (video call) or content that is broadcast live to a wider online audience. There is a substantial evidence base detailing the role that livestreaming plays in the commission of sexual exploitation of children, and offences relating to child sexual exploitation are discussed in chapter 6J: Unlawful immigration and human trafficking.

2B.60 Livestreaming can be used by perpetrators to evade detection because once the livestream is over, unless it was recorded, there may be little evidence of it. This risk has been noted by WeProtect, who have also highlighted how most platforms do not monitor livestreams.[450]

2B.61 The risk of livestreaming interacts with screen capturing and recording functionalities, as it allows for the livestreaming of CSEA to be used to create CSAM. The IWF's report on livestreaming analysed over 2,000 indecent images of children taken from livestreams and 98% of these showed children who appeared to be 13 years or under.[451]

2B.62 Livestreaming in conjunction with messaging functionalities could present added risks, as it allows perpetrators to make specific requests while an offence is taking place. The NSPCC found evidence that children who livestream are sometimes asked to perform sexual acts.[452] Of those children who livestream, 6% had received requests to change or remove their clothes.[453]

2B.63 As previously discussed, cases of livestreamed child sexual abuse have often been identified as being streamed from South-East Asia, with perpetrators in Western countries (including

---

[448] Westlake, B. G., & Bouchard, M., 2016. Liking and hyperlinking: Community detection in online child sexual exploitation networks. *Social science research*, 59, pp.23-36.

[449] Active Fence, 2023. How predators are abusing generative AI. [accessed 12 June 2024].

[450] In August 2022, the Australian e-Safety Commissioner issued the first mandatory transparency notices to Microsoft, Skype, Snap, Apple, Meta, WhatsApp, and Omegle, four of which have livestreaming or video call/conferencing services (note Omegle is no longer an active service). Responses revealed that of these four, three do not currently use tools to detect livestreamed child sexual abuse or exploitation: We Protect Global Alliance response to November 2023 Illegal Harms Consultation.

[451] IWF, 2018. Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse. [accessed 22 September 2023].

[452] Although not specified, it is reasonable to assume that these requests were received via messaging functionalities either publicly or privately, depending on the functionalities of the services being used and the tactics used by the perpetrator.

[453] 24% of all children have done a livestream broadcast: NSPCC, 2018. Livestreaming and video-chatting. [accessed 22 September 2023].

the UK) accessing the material, generally in exchange for payment. [454] Globally, the UK has previously been estimated to be the third largest consumer of this form of livestreamed child sexual abuse.[455]

2B.64 The livestreaming of child sexual abuse does not solely take place in South-East Asia. The IWF's research on livestreaming encountered many captures of livestreamed child sexual abuse which involved white girls, from apparently relatively affluent Western backgrounds, often appearing to be alone in their bedrooms.[456]

Direct messaging

2B.65 Direct messaging can allow perpetrators to share CSAM with one another. Interpol found that there was an increase in the volume of CSAM circulating via direct messaging on private messaging services or 'message applications' during the COVID-19 pandemic in 2020.[457]

2B.66 As outlined above, direct messaging is also a risk in the context of livestreaming, as it allows perpetrators to make specific requests while an offence is taking place.

Encrypted messaging

2B.67 Encrypted messaging makes the exchange of CSAM hard to detect,[458] which may increase the likelihood perpetrators will seek out spaces with encrypted messaging to disguise their activity. Protect Children reported that messaging apps are often favoured by offenders due to the perceived security and privacy offered by end-to-end encryption, which allows them to commit crimes with apparent reduced fear of detection or law enforcement presence.[459]

2B.68 Increased use of encrypted messaging may make the detection and reporting of CSAM more difficult. The IWF reported that Meta's suspected CSAM reports dropped by 58% between 2020 and 2021, during which time the IWF noted that Meta stopped 'voluntarily scanning' its services. The IWF has made the case that this scenario is similar to the situation where end-to-end encryption is rolled out and automated detection tools no longer work.[460] The NCA has estimated that most reports currently provided to UK police will be lost with the introduction of end-to-end encryption, impacting 92% of reports from Facebook and 85% from Instagram.[461]

2B.69 Evidence suggests increased use of end-to-end encryption may have a number of adverse impacts on child safety. Protect Children reported that it would hinder law enforcement efforts to identify and rescue victims and survivors and make it 'virtually impossible' to

---

[454] International Justice Mission, 2020. Online Sexual Exploitation of Children in the Philippines: Analysis and Recommendations for Governments, Industry, and Civil Society. [accessed 22 September 2023].
[455] IICSA, 2020. The Internet: Investigation Report. [accessed 22 September 2023].
[456] IWF, 2018. Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse. [accessed 23 August 2023]. [*Note: this research was funded by Microsoft].*
[457] Interpol, 2020. Threats and trends child sexual exploitation and abuse: COVID-19 impact. [accessed 22 September 2023].
[458] The exact scale of sharing and distribution of CSAM over encrypted messaging is difficult to quantify, as it cannot be tracked across services. Services offering end-to-end encryption have no means of accessing encrypted content. As such, technologies intended to mitigate the harm (such as hashing technology and content classifiers) cannot be applied within encrypted spaces and illegal content cannot be detected.
[459] Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].
[460] IWF, 2022. Not all Encryption is the same: social media is not ready for End-to-End Encryption. [accessed 30 August 2023].
[461] NCA, 2024. European police chiefs call for end-to-end encryption roll out to include public protection measures. [accessed 8 July 2024].

detect and remove CSAM, allowing for the cycle of revictimisation to continue. The report also noted it may contribute to the online disinhibition effect, as offenders become confident that their illegal activity cannot be detected.[462]

Ephemeral messaging

2B.70 Ephemeral messaging functionalities can make it harder for CSAM to be detected due to the disappearing nature of the content.

2B.71 Ephemeral messaging can be used by perpetrators to coerce children into producing and sharing sexual images, reassuring them that by using ephemeral messaging the image cannot be saved. However, it is common that perpetrators receiving such messages will screenshot the image.

2B.72 Ephemeral messaging has been found to facilitate SGII being shared consensually, as well as distributed non-consensually. Young people using these features may believe that their images are safer by sharing them in this format, in that there will be no permanent record of them, however the evidence suggests that users can deploy tactics to circumnavigate this feature.[463]

Group messaging

2B.73 Perpetrators tend to operate within groups or networks which trade content with each other. Group messaging functionalities can be used as part of these networks to share CSAM, enabling the spread of CSAM to multiple perpetrators. Research has found that abusers traded and shared CSAM in group chats on messaging applications.[464]

2B.74 Group messaging encompasses 'chat functions' which can be used by networks of perpetrators to share CSAM URLs with one another. Services may vary in their level of oversight or moderation of these channels, which presents a further risk if content sharing is not detected by human or automated moderation systems.

2B.75 Group messaging functionalities can be used to non-consensually share SGII to large groups of people, facilitating the quick spread of the material to a large number of users. Revealing Reality found that many young people on Snapchat are part of group chats of various sizes with their peers, such as school class, year group, friendship groups or groups based on extra-curricular clubs or social events.[465] Revealing Reality spoke with young people who had seen these images forwarded on to large group chats of their peers with just a couple of clicks.[466]

Commenting on content

2B.76 The ability to comment on content can be used to 'breadcrumb', whereby offenders use legal content to create a trail to direct like-minded individuals to illegal content, and thereby

---

[462] The report described how this may result in the emergence of large-scale CSAM communities: Protect Children (Lapsia, S.), 2024. Tech Platforms Used by Online Child Sexual Abuse Offenders. [accessed 20 June 2024].
[463] Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 22 September 2023].
[464] Steel, C., Newman, E., O'Rourke, S. and Quayle, E., 2020. An Integrative Review of Historical Technology and Countermeasure Usage Trends in Online Child Sexual Exploitation Material Offenders, *Forensic Science International: Digital Investigation*, Volume 33. [accessed 18 November 2024].
[465] Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 22 September 2023].
[466] Revealing Reality, 2022.

facilitate perpetrators' access to CSAM.[467] Some perpetrators use livestreaming comment functionality to start sexualised conversations with children and/or incite children to engage in sexual activity, sometimes including other children, in real time. The NSPCC found that of those children who livestreamed, 6% (more than 1 in 20) had received requests to change or remove their clothes.[468]

Posting content (text, images, video)

2B.77 The ability to post content, in this case text, videos and images, is a key enabler of the commission of CSAM offences. Abusers can post visual CSAM, and links or URLs to CSAM, on both open and closed channels of communication. The IWF reported that in 2022, 77% of the CSAM reports it dealt with were from services which hosted images.[469] These services typically allow users to post images which can subsequently be shared through a unique URL. Evidence indicates that such URLs are often 'embedded', presumably by being posted, on discussion forums (see Risk factor: service type section for more information).

Transactions and offers

Posting goods or services for sale

2B.78 The use of CSAM within an advert is a CSAM offence in itself, as well as enabling the commission of further CSAM offences.

2B.79 Perpetrators arranging the livestreaming of child sexual abuse for offenders to purchase may use online functions where goods and services are posted. Facilitators may include CSAM in their posts to advertise and attract offenders.

Online payments and cryptocurrency payments

2B.80 The ability to make online payments, as well as cryptocurrency payments, can enable CSAM offences. Cryptocurrencies or other exchange mechanisms, like vouchers or payment codes, can enable offenders to buy CSAM anonymously and evade detection. The IWF noted that the number of websites found to accept cryptocurrency payments for CSAM has doubled in most years since 2015.[470] Other potential payment options for CSAM include direct payment mechanisms, such as credit card or money transfer services.

Content exploring

Hyperlinking

2B.81 URLs, both in the form of hyperlinks and plain text, can be used by perpetrators to share CSAM between other individuals or more widely. Perpetrators can create links to CSAM stored on a file-storage and file-sharing service and share these across forums and in areas

[467] WIRED (Orphanides, K. G.) 2019. On YouTube, a network of paedophiles is hiding in plain sight. [accessed 22 September 2023].
[468] Survey of nearly 40,000 children aged 7 – 16 years old: NSPCC, 2018. *Livestreaming and video-chatting.* [accessed 22 September 2023].
[469] "*These sites provide 'storage' for images which either appear on dedicated websites or are shared within forums".* Source: IWF, 2023. The Annual Report 2022 #Behind the Screens: A deep dive into the digital and social emergency happening #BehindTheScreens, in children's bedrooms. [accessed 22 September 2023].
[470] IWF, 2022. Websites offering cryptocurrency payment for child sexual abuse images 'doubling every year'. [accessed 22 September 2023].

of otherwise legitimate services. Evidence has shown that large virtual communities of offenders have been seen to share millions of items of CSAM indirectly via hyperlinks.[471]

2B.82 Hyperlinks can facilitate access to CSAM as they can be used to direct abusers to CSAM hosted on third-party sites. The NSPCC refers to this as 'digital breadcrumbing', where abusers use services to signpost other abusers to CSAM hosted on other sites.[472] This includes the use of Quick Response (QR) codes, working in a similar way to hyperlinks.[473]

2B.83 The volume of link-sharing related to child sexual abuse also appears to be increasing, with offenders seemingly moving partly away from curating personal collections of CSAM and instead look to access 'on-demand' CSAM.[474]

2B.84 As noted previously, there is evidence of links to CSAM content being posted on social media sites in a 'scattergun' approach. The spamming of links to CSAM material drives up web traffic, and income, for those hosting CSAM content.[475]

Building lists or directories

2B.85 The ability to create lists or directories in folders and save them on file-sharing services can be used by perpetrators to collect particular kinds of CSAM. Using hyperlinks, perpetrators can then easily share their collections with other perpetrators.

User-generated content searching

2B.86 Autocomplete suggestions in a U2U service's search box can suggest searches for CSAM content. In 2018, Facebook's autocomplete search terms were found to suggest child abuse videos and 'under-age girls performing sex acts.'[476]

2B.87 In addition, some offenders use search functions on user-to-user pornography services to search for terms associated with CSAM.[477]

Content editing

Editing visual media

2B.88 As introduced previously, generative AI is being used by perpetrators to edit images and videos to produce CSAM. Perpetrators may use deepfakes, forms of audio-visual content that have been generated or manipulated using AI. Non-CSAM deepfakes are discussed in the Intimate image abuse chapter.

2B.89 The ability to edit images and videos can be used alter legal pornography, making participants appear as children, generating new CSAM material. Deepfake and GenAI technology be used to produce this type of CSAM, meaning that services allowing users to

---

[471] Westlake, B. G., & Bouchard, M., 2016. Liking and hyperlinking: Community detection in online child sexual exploitation networks. *Social science research*, 59, pp.23-36.
[472] NSPCC, 2022. Time to act: An assessment of the Online Safety Bill against the NSPCC's six tests for protecting children. [accessed 30 August 2023].
[473] NSPCC response to Ofcom 2022 Call for evidence: First phase of online safety regulation.
[474] WeProtect, 2023. Link-sharing and child sexual abuse: understanding the threat. [accessed 24 September 2024]
[475] IWF, 2022. Public warned as 'disturbing' new trend risks exposure to child sexual abuse material online. [accessed 30 August 2023].
[476] Hern, A., 2018. Facebook apologises for search suggestions of child abuse videos, The Guardian, 16 March. [accessed 30 August 2023].
[477] IWF, 2022. Internet Watch Foundation, Stop It Now, and Pornhub launch first of its kind chatbot to prevent child sexual abuse. [accessed 19 November 2024].

create such content through deepfake, GenAI or other content editing functionalities present risks.

Content storage and capture

Downloading content

2B.90 The ability to download content allows perpetrators to store and view local copies of CSAM on their computers and devices, as well as share it with others. This functionality enables perpetrators to build very large collections of CSAM. Indeed, the NCA has found some perpetrators who have downloaded over a million child sexual abuse images to their devices.[478]

Screen capturing or recording

2B.91 There is evidence of perpetrators capturing images from livestreams and capturing images of SGII videos from victims and survivors who have been groomed and coerced, which can then be distributed to other sites online.[479] Over a three-month period in 2017, the IWF found 2,082 child sexual abuse captures from livestreams online. Of these, 96% of the images were of children on their own, typically in a house; 98% of the images depicted children assessed as being 13-years-old or under; and 40% of the images were categorised as Category A or B. [480]

Recommender systems

Content recommender systems

2B.92 Recommender systems generally rely on user behaviour, such as viewing history, as an input into personalised content recommendations. As such, recommender systems are a risk factor in CSAM offences, as it is possible that a service's recommender system could suggest CSAM-related content to users who are actively viewing CSAM videos. In these instances, CSAM must be present in the content pool that the recommender system is sourcing, ranking, and serving content to users from.

2B.93 There is evidence of users being recommended inappropriate, but not necessarily illegal, content.[481] Salter and Hanson described how, if YouTube detects a user who seeks out and watches content of young children, the recommender system generates a playlist of similar content.[482]

---

[478] For example: Luck, F., 2023. Former GP caught with 1.2m indecent images of children jailed. BBC News, 23 February. [accessed 30 August 2023].

[479] While users can often record or capture content using third-party services, screen recordings and captures are shared on U2U services as user-generated content and some U2U services have dedicated screen recording and screen capturing functionalities.

[480] IWF, 2018. Trends in Online Child Sexual Exploitation: Examining the distribution of lives-streamed child sexual abuse. [accessed 22 September 2023]

[481] Fisher, M. and Taub, A., 2019. 'On YouTube's digital playground, an open gate for pedophiles', *New York Times*, 5 June. [accessed 27 September 2023].

[482] Salter, M. and Hanson, E., 2021. "I Need You All to Understand How Pervasive This Issue Is": User Efforts to Regulate Child Sexual Offending on Social Media. Chapter 42 in (Bailey, J., Flynn, A. and Henry, N.) Emerald International Handbook of Technology Facilitated Violence and Abuse, pp.729–748.

Risk factors: Business models and commercial profiles

Revenue models

Advertising-based model

2B.94 Advertising features on user-to-user services can be used by perpetrators of CSEA. Research by the IWF found that legitimate adverts can be inadvertently used to fund websites 'dedicated to child sexual abuse'[483] if those engaged in arranging the adverts (advertising agencies, brands and advertising exchanges) do not do enough to prevent their adverts' placement on such sites.[484]

Commercial profile

Low-capacity and early-stage services

2B.95 Low-capacity services, and services that are earlier in their business development lifecycle may be at greater risk of being used by perpetrators to share CSAM. Evidence suggests that low-capacity and early-stage services may be at risk of enabling CSAM content. This is because they are less likely to have technical and financial resources for risk management (for example, investment in the automated and/or manual moderation processes necessary to identify and combat CSAM content). We consider that perpetrators may seek out these spaces to share and view CSAM undetected.

Growth strategy

2B.96 Evidence suggests that services which prioritise and emphasise growth, for example prioritising user growth, may put insufficient resources towards effectively moderating harmful content and preventing offenders from exploiting the service.

[483] Home Office, 2018. Advertisers urged to help tackle online child sexual exploitation. [accessed 22 September 2023].
[484] The research found that "one in ten websites dedicated to child sexual abuse host adverts for legitimate brands, including some household names" and the preliminary research into a sample of child sexual exploitation websites found that 57 of 100 websites contained adverts: Home Office, 2018. Advertisers urged to help tackle online child sexual exploitation. [accessed 22 September 2023].

3. Hate

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for hate offences: how harm manifests online, and risk factors

Hate offences can be experienced by many people, in particular minorities and other protected groups. The offences can be targeted at one or more individuals, or wider communities. Exposure to hateful content, even if it may not meet legal thresholds, can have a severe impact on those to whom it is directed. The psychological effects of hateful content include shock, anger, suicidal thoughts, shame, exhaustion and fear, which can lead to further behavioural changes. Other experiences include financial harm and reputational damage. There is also evidence to suggest that, in some contexts, exposure to hateful content can entrench prejudices and incite acts of violence.

Service type risk factors:

Different types of services are associated with a risk of hate offences. **Social media services** and **online gaming services** pose a particular risk of hate offences. **Video-sharing services** and **private messaging services** have also been identified as spaces that are commonly used to commit or facilitate offences related to hate, targeting minorities and other protected groups.

User base risk factors:

Both large and small services play a role in disseminating hateful content. Users can first build a community of like-minded individuals to share provocative content on a **large** social media service, without breaching the service's terms and conditions for hateful conduct. They may then direct their user networks to **smaller and less-moderated services**.

Users' **race and ethnicity, gender, religion, disability status** and **sexual orientation** are all risk factors influencing someone's experience of hateful content. But age may also play a role.

Functionalities and recommender systems risk factors:

Evidence shows that the ability to create **anonymous user profiles** is a risk factor, but this is a complex issue. Anonymity can provide individuals with an environment in which they can speak and act more radically and propagate harm towards other users. But in many instances, hateful content is shared by users who are identifiable but unknown to the target. Usernames on identifiable **user profiles** can also be used to reference hate, while the ability to edit them can allow perpetrators to avoid enforcement action by recreating terminated profiles with slight edits to the original username.

Content recommender systems are another risk factor. Because these systems are generally designed to optimise user engagement, in some circumstances they promote content which may be hateful in nature because such negative or inflammatory content tends to have increased user engagement compared to benign content. Additionally, because recommender systems offer personalised feeds, they can also promote ideas or ideologies that users have already engaged with, which can increase confirmation bias and create 'filter bubbles'.

Hate offences are often committed via direct responses or **comments on posted content.** This functionality can enable the amplification of hate, enabling multiple people to direct abuse towards a target. The ability to **livestream** is also a risk factor as hateful content can be broadcast in real time. While **direct messaging** can be used to carry out hate offences in a highly targeted manner.

Other functionalities pose risks for hate offences. Evidence suggests that **content tagging,** using hashtags, is also a risk factor when services allow hashtags with hateful language to go unmoderated or unenforced. **User groups** can enable offenders to spread hateful content amongst like-minded users. **Hyperlinks** allow users to move easily from mainstream to more niche hateful spaces.

Business model risk factors:

Advertising-based revenue models with an incentive to maximise user engagement may sometimes advertently, or inadvertently, promote hateful content. However, advertisers can be sensitive to their adverts being associated with hateful content on a service and can use their economic leverage to require a service to protect against hateful content. Other revenue models may increase the risk in a different way; for instance, **subscription models** may be limited to a small group of like-minded users (subscribers) who share common views and so may be more tolerant of hateful content.

Introduction

3.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to the hate offences listed under 'Relevant offences' below; and
- The use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

3.2 We set out the characteristics of U2U services that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

3.3 We use the term 'hateful content' throughout this chapter to describe the content covered by these offences, which targets groups based on their race, religion or sexual orientation.

Research into online hate mainly focuses on hateful content that is prohibited by the terms and conditions of given services. Given this, the evidence described in this chapter includes sources that may cover non-statutory definitions relating to hate.

Relevant offences

3.4 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding hate offences, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

3.5 In this chapter, we consider the following public order offences that relate to stirring up racial hatred, religious hatred or hatred on the grounds of sexual orientation, specifically:

- The use of threatening, abusive or insulting words or behaviour, or the display of written threatening, abusive or insulting material intending or likely to stir up racial hatred.[485]
- The publication or distribution of threatening, abusive or insulting written material intending or likely to stir up racial hatred.[486]
- Distributing, showing or playing threatening, abusive or insulting recordings of visual images or sounds intending or likely to stir up racial hatred.[487]
- The use of threatening words or behaviour, or display of threatening written material intending to stir up religious hatred or hatred on the grounds of sexual orientation.[488]
- The publication or distribution of threatening written material intending to stir up religious hatred or hatred on the grounds of sexual orientation.[489]
- Distributing, showing or playing threatening, abusive or insulting recordings of visual sounds or images intending or likely to stir up religious hatred or hatred on the grounds of sexual orientation.[490]
- Racially or religiously aggravated harassment and public order offences.[491] This covers various offences relating to the fear or provocation of violence, harassment and stalking, when they are racially or religiously aggravated.

3.6 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and, in relation to offences in Scotland, being involved art and part in the commission of these offences).

3.7 The offences we are specifically considering in this chapter relate to race, religion and sexual orientation, in line with the priority offences set out in the Act. However, hateful content directed at other protected characteristics may also be illegal if, for example, it amounts to harassment or a public order offence. The Police and the Crown Prosecution Service (CPS) recognise hate crime based on race, religion, disability, sexual orientation, and

[485] Section 18 of the Public Order Act 1986.
[486] Section 19 of the Public Order Act 1986.
[487] Section 21 of the Public Order Act 1986.
[488] Section 29B of the Public Order Act 1986.
[489] Section 29C of the Public Order Act 1986.
[490] Section 29E of the Public Order Act 1986.
[491] Sections 31 and 32 of the Crime and Disorder Act 1998.

transgender identity. Crimes can be prosecuted as a hate crime if the offender has demonstrated, or been motivated by, hostility on the basis of these characteristics, and this may also have implications for sentencing.[492] We acknowledge that the experience of hateful offences is deeply personal and varies for each individual, and there may be a wider range of motivating factors not captured by the aforementioned definitions.

3.8 For this chapter, we focus on the specific offences set out above. However, in some places we also present evidence relating to hateful content directed at other protected characteristics where we consider that to be relevant. This chapter should also be read in conjunction with the Harassment, stalking, threats and abuse chapter.

3.9 Illegal hateful content can take many forms, including text, images, video and audio. Examples of such content online include the sharing of hateful messages that target minorities or protected groups. Social media services, as well as other open and closed online spaces, can provide a platform for broadcasting and disseminating hateful content, including inciting violence. The Harassment, stalking, threats and abuse chapter explores threats to kill in more detail. This may involve, or be done alongside, wider threats of violence to the individual.

3.10 For more details on the offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

How hate offences manifest online

3.11 This section is an overview which looks at how hate offences manifest online, and how UK individuals may be at risk of harm.

3.12 To put these risks of harm into context, Home Office research indicates that instances of online hate crime offences[493] towards a person, based on a protected characteristic[494] in the UK, are widespread. This research shows that 1,605 online hate crimes were recorded across 2017 and 2018 in England and Wales, representing about 2% of all hate crimes. The motivating factors for recorded online hate crimes were flagged as race (928), followed by sexual orientation (352), disability (225), religion (210) and transgender (69).[495] [496] However, we acknowledge these motivating factors are not exhaustive.[497]

3.13 Evidence suggests that a large proportion of people see or experience hateful content online. Ofcom research found that during a four-week period, 1 in 4 online users (adults and children aged 13-17) had seen or experienced content they considered to be hateful, offensive, or discriminatory, and which targeted a group or person based on specific

[492] The Crown Prosecution Service, n.d. Hate Crime. [accessed 7 June 2023].
[493] Refers to offences that have been recorded as hate crimes (flagged as being motivated by at least one of the five centrally monitored hate crime strands) and also been flagged as an online crime. We note that this is likely to be a broader definition than the specific priority offences that we are considering in this chapter.
[494] The Equality Act 2010 protects discrimination against someone with protected characteristics, which refers to: age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation.
[495] Home Office statistics on Online Hate Crimes were last given in 2017/2018, when experimental figures were reported for 30 out of 44 police forces.
[496] Home Office, 2018. Hate crime, England and Wales 2017/18. [accessed 2 April 2023].
[497] For instance, in Northern Ireland, hate crime data from the Police Service of Northern Ireland shows the prevalence of sectarian hate.

characteristics such as race, religion, disability, sexuality, or gender identity.[498] As a further, specific example, analysis conducted by the Woolf Institute estimates that on just one user-to-user service, 495,000 explicitly antisemitic posts are made viewable to UK users per year.[499]

3.14 Ofcom's research into hateful content online, conducted among a diverse sample of 39 people who had experienced online hate and hateful abuse, found that exposure to hateful content was a common feature of their online experience.[500] The frequency of hateful content experienced often increased after key national or international events such as a terror attack,[501] large sporting events like Euro 2020,[502] [503] [504] or more recently, events such as the Southport stabbing.[505]

3.15 TellMAMA, an anti-Muslim hate crime monitoring agency, confirmed a total of 1,109 online hate crime cases post the October 7th 2023 attacks on Israel by Hamas, which represented a more than threefold (335%) increase in anti-Muslim hate cases compared to the same time period a year earlier.[506] Hope not Hate, a UK-based advocacy group against racism and fascism, supports these findings. Through their research, they found there has been a significant increase in online anti-Muslim rhetoric in the UK following Hamas's attack on Israel.[507]

3.16 Online antisemitism is also on the rise, data from Community Security Trust, a British charity working against antisemitism and racism, recorded 630 cases of online antisemitism in the first half of 2024, which is an increase of 153% from the first six months of 2023, and the highest proportion of the total since 2020.[508]

3.17 Additionally, research by LGBTQ+[509] charity, Stonewall, found that one in ten LGBTQ+ people – including one in four trans[510] people – have been the direct target of homophobic, biphobic or transphobic abuse online.[511] Almost half of LGBTQ+ people (45%) have

---

[498] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 11 November 2024].
[499] Community Security Trust and Antisemitism Policy Trust, 2021. X: the extent and nature of antisemitism on X in the UK. [accessed 18 June 2024].
[500] Ofcom, 2023. Qualitative research into the impact of online hate. [accessed March 2023].
[501] Williams, M. and Reya, M 2019. Hatred behind the screens: a report on the rise of online hate speech. [accessed 23 March 2023].
[502] Ofcom, 2023. Qualitative research into the impact of online hate. [accessed March 2023].
[503] The Alan Turing Institute (Vidgen, B., Chung, Y-L, Johansson, P., Kirk, H.R., Williams, A., Hale, S.A., Margetts, H., Röttger, P. and Sprejer, L., 2022. Tracking abuse on X against football players in the 2021-22 Premier League season. [accessed 27 September 2022].
[504] Kearns, C, Sinclair, G, Black, J, Doidge, M, Fletcher, T, Kilvington, D, Liston, K, Lynn, T. and Rosati, P. A scoping review of research on online hate and sport. *Community and Sport.* 11 (2) [accessed 19 January 2023].
[505] Institute for Strategic Dialogue and CASM Technology, 2024. Evidencing a rise in anti-Muslim and anti-migrant online hate following the Southport attack. [accessed 15 October 2024]
[506] TellMAMA, Greatest Rise in Reported Anti-Muslim Hate Cases to Tell MAMA since Oct 7th. [accessed 22 October 2024]
[507] Hope not Hate, 2024. Doubling Down on Division, Anti Muslim hatred in the UK since 7th October. [accessed 22 October 2024]
[508] Community Security Trust, 2024. Antisemitic Incidents Report January-June 2024. [accessed 22 October 2024]
[509] LGBTQ+ is the acronym recognised by Stonewall.
[510] This term is recognised by the authors of the research to describe people whose gender is not the same as, or does not sit comfortably with, the sex they were assigned at birth – Stonewall, n.d. List of LGBTQ+ terms. [accessed 8 September 2023].
[511] Between February and April 2017, 5,375 lesbian, gay, bi and trans (LGBT) people across England, Scotland and Wales completed an online questionnaire about their life in Britain today. Source: Stonewall, 2017. LGBT in Britain: Hate crime and discrimination. [accessed 15 March 2023].

witnessed homophobic, biphobic or transphobic abuse or behaviour online that was directed at other people.

Risks of harm to individuals presented by the hate offences

3.18 Hateful content manifests in several ways online, including posts, memes, and comments on shared content. Ofcom research found that among respondents who had experienced hateful, offensive or discriminatory conduct online, nearly half (47%) came across it in comments on or replies to a post, article, or video; and the same proportion (47%) were exposed to it when scrolling through a service's feed or 'For You' page.[512]

3.19 The influence of demographic factors on risk is highly contextual and complex; many different demographic factors will be relevant to understanding the risks of harm to an individual and these intersections should be considered. Data indicates that user base characteristics including gender, sexual orientation, race and ethnicity, and religion lead to increased risk of individuals being targeted with hate content.[513]

3.20 Exposure to hateful content online can have a significant adverse impact on those to whom it is directed. Ofcom research indicates that the impact of online hateful content is more pronounced when the content targets a specific user or protected characteristic.[514]

3.21 The psychological effects of hateful content have been reported by those exposed as surprise and shock, anger and disappointment, embarrassment and shame, anxiety, fear, hopelessness and exhaustion. This can result in behavioural changes; anxiety and fear can lead to participants limiting what they share and express, or which online services they use.[515] Some users describe experiencing harm immediately after engaging with a one-off piece of content, while others described experiencing cumulative harm due to repeated exposure to potentially harmful content or interactions.[516]

3.22 Galop, a UK-based LGBT+[517] anti-violence charity, found that hateful content online against LGBT+ people led to people experiencing *"negative emotional responses to their online victimisation"* including fear, anxiety, self-blame and suicidal thoughts.[518]

3.23 Research has examined the link between hateful content online and hate crimes offline. Weak correlations were found between posts containing hateful language and specific types of crime. The strongest associations were found for religiously motivated crimes, but

[512] Respondents who reported having seen "Hateful, offensive or discriminatory content that targets a group or person based on specific characteristics like race, religion, disability, sexuality or gender identity; e.g. hate speech" online within the four weeks before completing the survey. Please note that this could include content which may not meet the threshold for illegal hate. Q. Which, if any, of the following have you seen or experienced online in the last 4 weeks? This includes any images, videos, audio or text, either comments, posts or messages you have seen and/or those shared directly to you. Source: Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 15 October 2024].
[513] See 'Risk factors: user base' section for more information.
[514] Ofcom, 2023. Qualitative research into the impact of online hate. [accessed March 2023].
[515] Ofcom, 2023.
[516] Ofcom, 2022. How people are harmed online: Testing a model from a user perspective. [accessed 2 November 2022].
[517] LGBT+ is the acronym recognised by Galop.
[518] Galop (Hubbard, L.), 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 12 September 2022].

not for racially motivated crimes. Although the results were inconsistent, they did *"point to the potential for using online behaviour to identify offline risk"*.[519]

3.24 Social science studies have observed how posts on social media services and other hateful content online can inspire acts of violence. The 2018 Pittsburgh synagogue shooter had posted antisemitic conspiracy theories on a small social media service, while perpetrators of white supremacist attacks have reportedly engaged with racist communities online.[520] It is important to note that exposure to hateful content online does not always correlate to online radicalisation, as a host of contextual factors interplay, often unique to individual circumstances.

Evidence of risk factors on user-to-user services

3.25 We consider that the risk factors we list are likely to increase the risks of harm relating to hate offences.

Risk factor: Service types

3.26 Research indicates that the following types of services can be a risk factor for hate offences: social media services, video-sharing services, private messaging services, and online gaming services.

Social media services

3.27 Social media services are a risk factor for hate offences targeting minorities and other protected groups.[521] Many social media services publish transparency reports[522] which include metrics for content removed when terms and conditions have been breached, which includes hateful content.[523] The publication of this data shows that social media services recognise that such content is present on their sites.

3.28 Our evidence shows that social media can be used to promote hateful ideologies and direct targeted hate against minority communities.[524] Additionally, by engaging with hateful content on social media services, some users subsequently form networks of like-minded individuals to proliferate hateful content online.[525] Once communities have been established, evidence indicates that users can be directed by the most active members to

[519] Cahill, M., Migacheve, K., Taylor, J., Williams, M., Burnap, P., Javed, A., Liu, H,. Lu, H. and Sutherland, A., 2019. Understanding online hate speech as a motivator and predictor of crime. [accessed 8 June 2023].
[520] Council on Foreign Relations, 2018. Hate speech on social media: global comparisons. [accessed 22 May 2023].
[521] See section above on Risks of harm to individuals presented by the hate offences for more information.
[522] With some exceptions, these are generally focused on content moderation metrics and government requests for user data and records. Source: Harling, A, Henesy D. and Simmance, E, 2023. View of Transparency Reporting: The UK Regulatory Perspective, *Journal of Online Trust & Safety,* 1(5). [accessed 22 September 2023].
[523] Further information on metrics can be found on individual platforms' published transparency reports.
[524] Institute for Strategic Dialogue (O'Connor, C.), Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. [accessed 22 October 2024].
[525] Poole, R, Giraud, E. and Quincey E, 2021. Tactical interventions in online hate speech: The case of #stopIslam, *New Media and Society,* 23(6) [accessed 12 Jan 2023].

less-moderated social media services where hateful ideologies can be discussed more openly.[526]

3.29 Social media services reliant on recommender systems may be particularly at risk of disseminating and amplifying hate offences. These systems are generally designed to optimise user engagement and can increase the risk of exposure to hateful content for users who have previously viewed similar content.[527]

Video-sharing services

3.30 Ofcom's own research has demonstrated the risk of content that amounts to hate offences being shared on video-sharing services is significant, with a third (32%) of people who use online video-sharing services reporting they had come across "hateful" content in the three months before the research took place.[528] This is despite significant amounts of potentially hateful content that violates services' community guidelines being removed each year.[529]

3.31 Research looking at specific events has also shown how comments on video-sharing services specifically, can be used to spread potentially hateful content. For example, ISD research in 2024 found that, in the wake of the Israel and Gaza conflict, anti-Muslim and antisemitic comments on posts published on a video sharing service rose sharply.[530]

Private messaging services

3.32 Private messaging services can be used to create inward-looking groups, which can be perceived as a safe space to stir up hatred based on race or ethnicity, religion or sexual orientation. This is particularly true of services with end-to-end encryption, due to the added security and privacy they offer users and the subsequent challenges that this presents to the detection and moderation of harmful content. However, these services can also be used to disseminate hateful narratives and therefore be used to commit or facilitate hate offences. Research conducted by the Anti-Defamation League (ADL) assessed some of the online content present on Telegram in the USA and found that it comprised a large proportion of references to Jewish and Black people.[531] Of the 333,325 Telegram messages analysed, one in every 81 messages was derogatory towards Black people in America, and about one in every 54 messages were derogatory towards Jewish Americans.[532]

---

[526] N, Velasquez., R, Leahy., N, Johnson Restrepo., Y, Lupu., R, Sear., N, Gabriel., O, K, Jha., B Goldberg. and N, F, Johnson., 2021. Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms. *Scientific Reports*, 11 (11549). [accessed 20 February 2023].

[527] Reed, A, Whittaker, J, Votta, F. and Looney, S., 2019. Radical filter bubbles: social media personalisation algorithms and extremist content. [accessed 19 February 2023].

[528] A third of users (32%) said they had witnessed or experienced hateful content. Hateful content was most often directed towards a racial group (59%), followed by religious groups (28%), transgender people (25%) and those of a particular sexual orientation (23%). Source: Ofcom and Yonder, 2021. User Experience of Potential Online Harms within Video Sharing Platforms. [accessed 20 September 2024].

[529] For example, between April and June 2024 Google removed approximately 34,000 channels for "Hateful or abusive" violations of their community policy. Note that this accounted for only 1% of all channel removals in this time period. Source: Google, 2024. Transparency report. [accessed 20 September 2024].

[530] Institute of Strategic Dialogue, 2023. 43-fold increase in anti-Muslim YouTube comments following Hamas' October 7 attack. Institute of Strategic Dialogue, 2023. Rise in antisemitism on both mainstream and fringe social media platforms following Hamas' terrorist attack. [accessed 10th May 2024] Narratives of Hate: Post-7 October Antisemitism and Anti-Muslim Hate on Social Media. [accessed 18th June 2024].

[531] ADL (Kumbleben, M., Woolley, S. and Engler, M.), 2020. Computational propaganda and the 2020 U.S. Presidential election: Antisemitic and anti-Black content on Facebook and Telegram.[accessed 2 November 2022].

[532]ADL (Kumbleben, M., Woolley, S. and Engler, M.), 2020.

3.33 Ofcom research revealed how hateful content can be shared via private messaging services: one case study demonstrated how a non-binary individual was subjected to a 'hate raid' on a private messaging service linked to a livestream. During a celebratory livestream on a social media service, the chat on the private messaging service became flooded with 'bots' and 'raiders' whose usernames and profile pictures contained racist slurs and imagery. The individual felt they had been targeted due to their LGBTQ+ status.[533]

Online gaming services

3.34 Online gaming services can also be used to spread and enable hateful content. A United Nations report examining gaming and violent extremism noted that certain gaming communities facilitate *"a culture in which misogyny, toxicity, racism and hate can flourish"*.[534] An investigation by the BBC concluded that "*extremists are using mainstream video games and gaming chat platforms to spread hate*", with researchers finding "*extremist roleplay scenarios within games on various platforms*" that included "*Nazi concentration camps and a Uyghur detainment camp*".[535] This research also found "*antisemitism, racism and homophobia on platforms [where] users stream and chat about games*".[536]

Risk factors: User base

User base size

3.35 The number of users on a service carries different risks associated with hateful content. Services with a large user base, as well as smaller, niche services, can be at risk. However, there is evidence that niche online services can contain far more abuse, including hateful activity, than mainstream services, despite these services attracting far fewer users.[537]

3.36 Perpetrators can take advantage of these differences in services, using them as ecosystems to fulfil their motivations. Perpetrators of hate offences tend to use services with large and small user bases in different ways. Research has found that some potential perpetrators are incentivised to maintain a presence on larger mainstream social media services, where they build their network further with new users, attracting them with 'borderline' hateful content, such as by sharing incendiary news stories and provocative memes. These networks are then directed towards less-moderated services. In these spaces, users discuss and share hateful content more openly.[538] It is possible that these less-moderated services have smaller user base sizes.

---

[533] Ofcom, 2022. How people are harmed online: Testing a model from a user perspective. [accessed 2 November 2022].
[534] Research included an online survey of gamers, n=622, focus groups with six avid gamers and focus groups with six experts. United Nations Office of Counter-Terrorism, 2022. Examining the Intersection Between Gaming and Violent Extremism. [accessed 3 June 2023].
[535] Miller, C. and Silva, S., 2021. Extremists using video-game chats to spread hate, *BBC,* 23 September. [accessed 22 September 2023]..
[536] Miller, C. and Silva, S., 2021.
[537] The Alan Turing Institute (Vidgen, B., Margetts, H. and Harris, A.), 2019. How much online abuse is there? A systematic review of evidence for the UK. [accessed 12 July 2023].
[538] Velasquez, N., Leahy, R., Johnson Restrepo, N., Lupu, Y., Sear, R., Gabriel, N., Jha, O. K., Goldberg, B., and Johnson, N. F., 2021. Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms, *Scientific Reports*, 11 (11549) [accessed 20 February 2023].

User base demographics

3.37 The following section outlines the primary evidence on user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

3.38 Data suggests that user base characteristics including age, gender, sexual orientation, race and ethnicity, and religion affect the risks of harm to individuals.

3.39 Users' likelihood of being exposed to hateful content online appears to be correlated with their age. Ofcom research found that on average, 25% (1 in 4) of UK internet users had seen or experienced "*hateful, offensive, or discriminatory content that targeted a group or person based on specific characteristics like race, religion, disability, sexuality, or gender identity, (hate speech)"* in the four weeks leading up to the study.[539] 35% (more than 1 in 3) of users aged 18 to 24 reported seeing such content, compared to 25% of those aged 45 to 54 and 18% of those aged 55 to 64.[540] The Oxford Internet Survey of internet use in Great Britain[541] also found that younger people are more likely to experience abuse online; four in 10 (41.2%) of 18 to 30-year-olds had seen cruel or hateful content online, compared with 7.4% of over-75s.[542]

3.40 The ethnicity of a user also affects how likely it is that they will be exposed to hateful content. Ofcom research found that 33% of UK internet users who defined themselves as any mixed ethnicity had seen or experienced *hateful, offensive, or discriminatory content* online in the past four weeks, compared to 24% who defined themselves as white.[543]

3.41 The Oxford Internet Survey also found that ethnicity impacted respondents' experience of online abuse; 26.6% of white respondents had viewed cruel/hateful content online compared to 38.6% of Black respondents.[544]

3.42 Black women are likely to be exposed to racist and misogynistic content online according to research by Glitch which found over two thousand "highly toxic"[545] posts across five social media platforms. Content sat at the intersection of racist and sexist hate, with content about Black women ranking as the most toxic on average, including the use of dehumanising language and stereotypes. 69.9% of the posts labelled 'Black toxic' were in the highest category of toxicity, compared to 33.6% of the 'white toxic' data set.[546]

3.43 The report by Glitch demonstrates the role that intersectionality[547] can play in heightening the risk of certain groups of people encountering hate online – where individuals with multiple personal characteristics may be targeted more often and with more severe forms

---

[539] Ofcom, 2024. Online Experiences Tracker - Wave 6. [accessed 11 November 2024].
[540] Some of the content seen by users may not meet the legal threshold for the relevant offences we are considering in this chapter. Source: Ofcom, 2022. Online Experiences Tracker Data tables waves 1 and 2. [accessed 18 July 2023].
[541] Multi-stage national probability sample of 2,000 people.
[542] The Alan Turing Institute (Vidgen, B, Margetts. and H, Harris, A.), 2019. How much online abuse is there? A systematic review of evidence for the UK. [accessed 12 July 2023].
[543] Ofcom, 2024. Online Experiences Tracker - Wave 6. [accessed 11 November 2024].
[544] The Alan Turing Institute (Vidgen, B., Margetts, H. and Harris, A.), 2019.
[545] Keywords for toxic messages include racial and misogynistic slurs.
[546] Glitch, 2023. The Digital Misogynoir Report: Ending the dehumanising of Black women on social media, p.35. [accessed 18th June 2024]
[547] A term created by American sociologist Kimberlé Crenshaw to describe how people can face different kinds of discrimination at the same time due to their "intersecting" or overlapping personal characteristics.

of hateful abuse. For instance, Muslim women and Jewish women may be more susceptible to being targeted with hate due to their religious and gender characteristics, much like how Black women would be due to their racial and gender characteristics. Since intersectional hate manifests through the coming together of an individual's or community's personal characteristics (such as race, religion, gender, and class) it creates a unique and more complex type of discrimination.

3.44 That is why religion is another characteristic which we are assessing in relation to the stirring-up of hatred. A user's religion can be a risk factor in the exposure of hateful content. Ofcom research found that of the 22% (nearly 1 in 4) of UK internet users who had seen or experienced hateful, offensive, or discriminatory content online in four weeks prior to responding to our survey, Jewish internet users (51%) (more than half) and Muslim internet users (33%) (1 in 3) were more likely to report having seen or experienced such content than, for example, Church of England/Scotland/Ireland users (15%).[548]

3.45 Individuals belonging to the disabled community are also more vulnerable to being targets of online hate. Disability Equality charity SCOPE's research on online trolling revealed that one in ten disabled people have experienced online abuse. More than half (53%) have seen negative comments about disabled people or disability in general.

3.46 In a meeting with The Coventry Youth Activists (CYA), CYA told Ofcom that they have faced a lot of disability-based hate on their personal profiles.[549] Their experiences support SCOPE's study which showed that younger people are the most vulnerable within the disabled community with almost half of 18- to 34-year-olds enduring negative comments online.

3.47 The sexual orientation of users is also a risk factor in exposure to hateful content online. Ofcom research found that LGBTQ+ UK internet users were significantly more likely to report having seen or experienced hateful, offensive or discriminatory content online in the past four weeks. Gay/lesbian (38%) (nearly 4 in 10) and bisexual UK internet users (56%) (more than half) were more likely to report having seen or experienced such content than heterosexual users (27%).[550]

3.48 Galop's Online Hate Crime Report surveyed over 1,100 LGBT+ people and found that 60% (6 in 10) of respondents had experienced anti-LGBT+ abuse online, and trans individuals are more likely to experience online abuse than their cis-gendered counterparts.[551] Therefore, being transgender is another risk factor for online hate.

3.49 The Anti-Defamation League (ADL) published research in 2020 into online hate and harassment.[552] Its findings indicated that LGBT individuals, Muslims, Hispanics or Latinos, and African Americans faced particularly high rates of identity-based discrimination. In comparison to the same survey the previous year, respondents reported a doubling of

[548] Ofcom, 2024. Online Experiences Tracker - Wave 6. [accessed 11 November 2024].
[549] Ofcom / Coventry Youth Activists meeting, 16 April 2024.
[550] Ofcom, 2024. Online Experiences Tracker - Wave 6. [accessed 11 November 2024] – 18+ only, of the 26% who has seen hateful, offensive or discriminatory content
[551] Galop (Hubbard, L.), 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 12 September 2022].
[552] Nationally representative survey of Americans, 1,974 respondents: ADL, 2020. Online hate and harassment report: The American experience 2020. [accessed 20 October 2022].

religion-based harassment (from 11% to 22%) while race-based harassment had increased from 15% to 25%.

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles

3.50 The ability to share content anonymously, and the impact this has on the risk of hateful content occurring on a service, presents a complex picture. While some studies suggest that anonymity can increase the risk of users sharing hateful content, a significant amount of hateful content is shared by users who are not anonymous.

3.51 Some users have experienced hateful conduct online posted anonymously. Galop's Online Hate Crime Report 2020 reported that 20% of online anti-LGBT+ hate incidents were committed by anonymous users.[553] However, within the same study a significant number of online hate incidents were perpetrated by an offender 'not anonymous, but unknown to' the user (52%). This study shows that not all the hate incidents researched were committed by anonymous users, but the perpetrators were often not known to the targets.

3.52 Mondal *et al.* conducted research into the use of explicit hate expressions in social media services. They noted that anonymity plays an important part in contributing to polarising discussions. The research found that hateful content on a given social media service concerning race or sexual orientation was more likely to be posted anonymously.[554] Research into the link between anonymity and abuse online has also highlighted the role pseudonymous user profiles play. Users even described creating 'disposable' user accounts in order to carry out abuse online in a particular case or for a more limited time, knowing their actions were in contravention of a service rules.[555]

3.53 However, we recognise that anonymity can be important in protecting users and allowing people to express themselves and engage freely online; for example, users who wish to talk openly about their sexuality or explore gender identity without fear of discrimination or harassment.[556] Anonymity can enable users to express ideas or criticisms about people in power without risking attribution.[557] Digital rights campaigners have argued that anonymity is a "*crucial tool for women and sexual minorities. The use of anonymity online supports the most vulnerable groups*".[558] Therefore, while anonymity online may give rise to some risks, it also confers some important benefits.

User profiles

3.54 Usernames on user profiles can be used to stir up hatred against groups with protected characteristics, by intentionally publishing or distributing threatening material. Evidence

[553] Galop (Hubbart, L.), 2020.
[554] Mondal, M, Silva, L. and Benevenuto, F., 2018. Characterising usage of explicit hate expressions in social media, *New Review of Hypermedia and Multimedia,* 24(2). [accessed 18 November 2022].
[555] Revealing Reality, on behalf of the Department for Media, Culture and Sport. 2022. Abuse and Anonymity. [accessed 23 October 2024].
[556] eSafety Commissioner, n.d. Anonymity and identity shielding [accessed 22 May 2023].
[557] eSafety Commissioner, n.d.
[558] eSafety Commissioner, n.d.

suggests that usernames have been used as a tool by users to spread racial slurs.[559] Users have also returned to a service after enforcement action by slightly editing their username. The Institute for Strategic Dialogue (ISD) found that accounts which had been banned from TikTok sometimes returned to the service under an edited username.[560] Users can often easily find the banned user via their username, even after the original account has been banned.[561]

User communications

Livestreaming

3.55 Evidence suggests that livestreaming can be a risk factor for hate offences due to the ease with which a user can reach a large audience quickly. As evidenced in other chapters, such as Terrorism, Child sexual exploitation and abuse (CSEA), and Extreme pornography, livestreaming can be used to broadcast illegal content in real time.

3.56 Livestreams can also be used to share hateful content. For instance, leading up to the Southport riots in July 2024, the Online Safety Act Network's (OSAN) analysis pointed to the use of livestreaming to incite hate and violence across the country.[562] Thus, livestreams can be used to broadcast hateful content with large audiences of users, and their ephemeral nature makes moderation challenging.[563] Moreover, some video-sharing services allow users to combine user-generated content (UGC) with existing content, which can be used to respond to posts in a hateful way.[564] The ephemeral nature of livestreaming means that the content is less likely to be archived and may not be moderated in real time. [565] The risk of harm presented by livestreaming is increased when paired with screen recording functionality, as the subsequent recording and dissemination of potentially hateful livestreamed footage can increase content virality.[566]

Direct messaging

3.57 As discussed in the chapter 'Harassment, stalking, threats and abuse', evidence shows that direct messaging can be used to carry out these harms in a targeted manner. This type of conduct could be hateful if the messages are racially or religiously aggravated; Ofcom research found evidence of direct messaging being used by perpetrators to target a victim with racist abuse.[567]

Commenting on content

3.58 Some services allow users to reply to or comment on posted content, and the volume of such comments can increase in response to external events. Real-world external events can

---

[559] Institute for Strategic Dialogue (O'Connor, C.), 2021. Hatescape: An in-depth analysis of extremism and hate speech on TikTok. [accessed 9 March 2023].
[560] Institute for Strategic Dialogue (O'Connor, C.), 2021.
[561] Institute for Strategic Dialogue (O'Connor, C.), 2021.
[562] Online Safety Act Network, 2024. Disinformation and disorder: the limits of the Online Safety Act. [accessed 30 October 2024].
[563] Zhou, Y. and Farzan, R., 2021. Designing to stop live streaming cyberbullying: A case study of Twitch live streaming platform. [accessed 22 September 2023].
[564] Institute for Strategic Dialogue (O'Connor, C.), Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. [accessed 22 October 2024].
[565] Zhou, Y. and Farzan, R., 2021. Designing to stop live streaming cyberbullying: A case study of Twitch live streaming platform. [accessed 22 September 2023].
[566] Ofcom, 2022. The Buffalo Attack: Implications for Online Safety. [accessed 18 January 2023].
[567] Ofcom, 2023. Qualitative research into the impact of online hate. [accessed 22 September 2023].

trigger an increase in online hate speech facilitated by commenting functionalities. For example, ISD research found that in the wake of the Israel and Gaza conflict, anti-Muslim and antisemitic comments on posts published on a video sharing service rose sharply.[568]

3.59 Because many different people can post comments at one time, targets of hate can receive multiple attacks at once or in a very short space of time, potentially amplifying the risk of harm through the cumulative impact of each attack. Ofcom research into online abuse received by footballers suggested that while users may send just one abusive comment to an individual,[569] sometimes the targeted individual would receive comments from many users simultaneously. Galop research into online hate crime against LGBT+ individuals found that incidents were likely to involve more than one perpetrator, with 71% of online anti-LGBT+ incidents involving more than one perpetrator and 13% of incidents involving more than 20 perpetrators. This has led to respondents reporting incidents of 'cybermobbing' and/or 'dogpiling'.[570] [571]

Reacting to content

3.60 Users' reaction to content is an important part of the system in which people who post abusive content online do so, and are further encouraged by, the status that it brings them, particularly within communities of like-minded individuals.[572] Research examining the 'Stop Islam' hashtag on X, following the 2016 Brussels terror attack, found that users who engaged with hashtags, or liked or retweeted content about the event, could form a tightly bound network of like-minded individuals. This could result in close-knit connections between users, enabling the rapid spread of content such as memes that parody counter-speech to hateful content by opposing groups.[573]

Posting content

3.61 The ability to post content on services enables its easy dissemination, increasing the risk of exposure of users to controversial or emotive posts, and making it easier to disseminate hateful content. In the ten days following the Southport attack, ISD found that the use of anti-Muslim slurs in posts more than doubled on X (formerly Twitter).[574] Hashtags containing anti-Muslim sentiments also proliferated, collectively receiving almost five million views. ISD's findings supports previous research which shows that posts featuring

---

[568] Institute of Strategic Dialogue, 2023. 43-fold increase in anti-Muslim YouTube comments following Hamas' October 7 attack. [accessed 15 October 2024]; Institute of Strategic Dialogue, 2023. Rise in antisemitism on both mainstream and fringe social media platforms following Hamas' terrorist attack. [accessed 10th May 2024]; Institute of Strategic Dialogue, 2024. Narratives of Hate: Post-7 October Antisemitism and Anti-Muslim Hate on Social Media. [accessed 18th June 2024].

[569] Ofcom's research into X abuse of Premier League football players found that many users send just one abusive tweet. Source: The Alan Turing Institute, (Vidgen, B., Chung, Y-L, Johansson, P., Kirk, H.R., Williams, A., Hale, S.A., Margetts, H., Röttger, P. and Sprejer, L.), 2022. Tracking abuse on X against football players in the 2021-22 Premier League season. [Accessed 27 September 2022].

[570] Galop (Hubbard, L.), 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 12 September 2022].

[571] Both terms refer to more than one person directing abusive comments towards an individual in a semi-co-ordinated manner.

[572] Revealing Reality, on behalf of the Department for Media, Culture and Sport. 2022.

[573] Poole, R., Giraud, E., and Quincey E., 2021. Tactical interventions in online hate speech: The case of #stopIslam, *New Media and Society,* 23(6). [accessed 12 Jan 2023].

[574] Institute for Strategic Dialogue and CASM Technology, 2024. Evidencing a rise in anti-Muslim and anti-migrant online hate following the Southport attack. [accessed 30 Sep 2024].

hashtags and containing antisemitic comments can reach large audiences on these services.[575]

3.62 Evidence also indicates that bots, which could include those employing GenAI technologies, can operate user accounts that post hateful content on services. The use of bots increases the volume of hateful content on a given service, passively exposing more users to hateful bot-generated content. Peer-reviewed research focusing on bot-activity and hate speech on X found that six months into the COVID-19 pandemic, hateful content was largely being produced by bots.[576] The volume of bots indicates that some services may have insufficiently robust measures to prevent the use of bots and are therefore likely to pose a higher risk of exposure to hate speech.

User networking

User groups

3.63 Services that allow users to build online communities with one another are a risk factor, as they can enable offenders to spread hateful content among like-minded users, in an environment where users will encourage each other to do and provide a sense of status that can be derived, in part, from saying or posting content that is particularly or increasingly extreme.[577] In a study exploring the experiences of former members of violent and racist groups, a former member discussed how online communities enabled members to spread hateful content which "*facilitated the process of violent radicalisation*".[578] This study provides anecdotal evidence in relation to how potential perpetrators view online communities as an opportunity to spread and disseminate hateful content.

User-generated content exploring

Hyperlinking

3.64 Hyperlinking can be a risk factor to distribute hateful content. For example, research by the Data and Society Research Institute into the 'AIN' (Alternative Influencers' Network) found that with the help of guest appearances[579] and other hyperlinks, audiences on a video-streaming service can easily move from mainstream to hateful content.[580]

Content tagging

3.65 Content tagging can be a risk factor when users promote hateful content, thereby heightening the risk of its visibility and user exposure. This includes efforts – which may or may not be successful – to use the algorithmic function of certain hashtags to achieve views and engagement. Research by ISD revealed that users who post hateful content make use of popular hashtags on video-streaming services. They also use hashtags relating to general political discussion and trends, indicating that they assume that the algorithmic systems of

---

[575] CCDH, 2021. Failure to protect: How tech giants fail to act on user reports of antisemitism. [accessed 20 September 2023].
[576] Uyheng, J., Bellutta, D. and Carley, K., 2022. Bots amplify and redirect hate speech in online discourse about racism during the Covid-19 pandemic. *Social Media + Society,* 8(3). [accessed 4 September 2022].
[577] Revealing Reality, on behalf of the Department for Media, Culture and Sport. 2022.
[578] Gaudette, T., Scrivens, R. and Venkatesh, V., 2020. The role of internet in facilitating violent extremism: Insights from former right-wing extremists. *Terrorism and Political Violence.* 7 (34). [accessed 22 September 2023].
[579] A guest appearance refers to a person collaborating with a service user and taking part in the video. They can often be notable online personalities or 'influencers'.
[580] Data & Society (Lewis, R.), 2018. Alternative influence: broadcasting the reactionary right on Youtube. [accessed 20 February 2023].

services will promote certain trending topics to wider audiences.[581] Some users promoting hate know how to exploit algorithms, increasing the risks of harm to other users, as content they would not normally search for is likely to appear on their feeds under 'trending' topics.[582]

3.66 Research by the charity Antisemitism Policy Trust found that antisemitism is present on some social media services, with antisemitic hashtags often associated with conspiracy theories. Such hashtags are sometimes attached to posts that have no direct relationship to the content of the post, causing them to be displayed to a large pool of users, who in most cases are not actively looking for antisemitic content but are unwittingly exposed to it. The research reports that antisemitic hashtags, alongside hashtags with demonstrable links to antisemitism, were viewed on a social media service tens of thousands of times during a seven-week period, and generated thousands of likes in response.[583] A high-profile example of this hateful content was Luciana Berger, the former MP, who was subjected to antisemitic harassment, with the perpetrator using antisemitic hashtags on a social media service as part of his campaign of abuse.[584] Antisemitic hashtags and their relationship to conspiracy theories is an example of religious discrimination online, and there are risks of harm to those exposed to this content.

User-generated content editing

Editing visual media

3.67 Although limited, our evidence suggests that the ability to edit user-generated content can be used to commit or facilitate hate offences. For example, some video-sharing services allow users to combine user-generated content with existing content. These features can be exploited by users to post hateful content in response to existing creators' posts.[585] Evidence has also shown that hateful messages can be widely disseminated using edited popular memes or through the creation of striking imagery with hateful content overlapping it. This content may glorify attacks targeting specific minority ethnic groups or religious affiliations and can encourage similar attacks to continue.[586]

Recommender systems

Content recommender systems

3.68 As recommender systems are designed to curate personalised content feeds, they can make it more likely that users who engage with hateful content see more of it in the future. Hateful content may also be recommended across user accounts which have shared engagement patterns. If users are engaging sufficiently with hateful content the

[581] Institute for Strategic Dialogue (O'Connor, C.), Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. [accessed 22 October 2024].
[582] This relates to services which employ trending functionalities.
[583] Antisemitism Policy Trust, 2021. Instagram: Bad Influence. [accessed 20 December 2022].
[584] BBC News, 2016. Man jailed for harassing Labour MP Luciana Berger. BBC News, 8 December. [accessed 22 September 2023].
[585] Institute for Strategic Dialogue (O'Connor, C.), Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. [accessed 22 October 2024].
[586] CST, 2020. Hate fuel: The hidden online world fuelling far right terror. 11 June. [accessed 7 June 2023].

recommender system may then promote this content, which could be divisive, untrue, or incendiary.[587]

3.69 The mechanism that underpins the dissemination of hate content is driven by explicit user feedback such as likes, comments, and shares (positive engagement). Where users express interest in content through positive user feedback, recommender systems are likely to amplify that content. Consequently, if a user is primarily engaging with hateful content and not with other types of content, this is then more likely to create a 'filter bubble', where the user is recommended more hate content while other content is deprioritised.

3.70 A filter bubble is where a user is recommended items that reinforce their own preferences. As a result, recommender systems can facilitate confirmation bias. A study into a large video-streaming service found that a user account which predominantly interacted with hateful content was twice as likely to be shown more extreme hateful content, and more likely to be recommended 'fringe' content.[588] This suggests that when users interact with hateful content on the platform, it is further amplified to them in the future. However, this was not the case on two other services that were studied.[589]

3.71 Munn *et al.* also found that recommender system design on a social media service can stimulate the user with outrage-inducing content while enabling seamless sharing, allowing content to rapidly proliferate across the network. This increases the prevalence of such content, making it easier for users to discover and engage with.[590]

Risk factors: Business models and commercial profile

Revenue models

Advertising-based revenue model

3.72 Advertising-based revenue models with an incentive to maximise user engagement and time spent may sometimes advertently, or inadvertently, promote hateful content if this increases engagement. However, advertisers can be sensitive to their adverts being associated with hateful content on a service and can use their economic leverage to require a service to protect against hateful content.

3.73 Services for which advertising is a primary income stream are incentivised to report to advertisers a high user base and high user time spent, as these are key to attracting advertisers to the service. Therefore, services which rely on advertising revenue models have a financial incentive to promote content that drives user engagement.

3.74 This may – intentionally or unintentionally – promote hate speech activity if it increases user engagement and attracts advertising revenue, particularly in instances where content moderation systems fail to detect hateful content. For example, the article *Angry by design:*

[587] Munn, L., 2020. Angry by design: Toxic communication and technical architectures. *Humanities and Social Sciences communications*. 7 (53) [accessed 3 May 2023].
[588] "Fringe" content was coded as radical content without justification of violence, may also include profanity laden nicknames that go beyond political discourse, or historical revisionism.
[589] Whittaker, J., Looney, S., Reed, A., Votta, Fabio., 2021. Recommender systems and the amplification of extremist content. [accessed 10 October 2024]
[590] Munn, L, 2020.

toxic communication and technical architectures[591] sets out the argument that social media services which rely on advertising revenues try to increase these revenues through increasing user interaction with the platform. An effective way of doing this is to display, and facilitate, a large quantity of controversial content, which could include hateful content. Material that creates outrage or strong reactions can encourage user engagement and by extension, increase the amount of time spent viewing the material.

3.75 However, we recognise that this risk will depend on the extent to which the service faces commercial pressure from advertisers (or indirectly, from users) to remove hate speech activity. Advertisers' reactions to their adverts being placed on a service which may host online harm activity can be a primary driver in the service's attitude to tackling online harms. For example, two initiatives show that advertisers can act to protect against harms such as hate content. Such advertisers are active in organisations whose goals are *"eliminating harmful online content and ensuring that bad actors have no access to advertiser funding"*,[592] and *"to break the economic link between advertising and the harmful content".*[593]

Subscription-based revenue models

3.76 Other revenue models may increase the risk in a different way; for instance, subscription models may be limited to a small group of like-minded users (subscribers) who share common views and so may be more tolerant of hateful content.

1.86 Services on which hateful content appears may also be supported by non-advertising income streams such as subscriptions or donations. Subscribers and donors may identify with the values of the service, and the hateful content on it, creating a self-enclosed community where a variety of opinions is unavailable and hateful content is 'normalised', with the income streams contributing to the ongoing provision of the service.[594]

[591] Munn, L., 2020. Angry by design: Toxic communication and technical architectures. *Humanities and Social Sciences communications*. 7 (53) [accessed 3 May 2023].
[592] World Federation of Advertisers, 2020. Marketing leaders take action on harmful online content. [accessed 22 September 2023].
[593] Conscious Advertising Network, n.d. About Us. [accessed 22 September 2023].
[594] Stanford (Thiel, D. and Mcain, M.), 2022. Gabufacturing dissent: An in-depth analysis of Gab. [accessed 3 May 2022].

4. Harassment, stalking, threats and abuse

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for harassment, abuse, threats, stalking and threatening communications offences: How harms manifest online, and risk factors

This chapter covers offences relating to online harassment, abuse, threats, stalking or threatening communications which are unwanted behaviours that can cause alarm and distress to other individuals, or put them in fear of violence. Ofcom's Online Experiences Tracker showed that 12% of people aged 13+ surveyed, reported they had witnessed 'one off abusive behaviour or threats in the past four weeks.595 Additionally, the Oxford Internet Survey in 2019 reported that 27% of people had been exposed to abusive content. 596 They can cause significant harm to individuals. Psychological impacts can include mental and emotional distress, isolation, and feeling unsafe both online and offline.

Service type risk factors:

Research indicates that **social media services** are used to commit and facilitate various forms of abuse and harassment, including threats to kill, and misleading information that can result in violence.

Evidence also indicates that users regularly experience severe abuse, including physical threats, stalking, and sustained harassment, on **online gaming services.**

These offences are also perpetrated on **private messaging services** and **online dating services**.

Userbase risk factors:

Evidence suggests that anyone can be subjected to these behaviours. However, research shows that women are particularly subjected to harassment. In a study surveying 4500 women globally, 38% of women aged 18 to 74 stated that they had personal experiences of "online violence".[597] In Plan International's study of over 14,000 girls and women aged 15 to 25, 58% of respondents reported they experienced some form of online harassment.[598] Their exposure to these offences is often more prevalent, severe, and can cause greater effect, compared to men,

[595] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].
[596] 27% of respondents had seen cruel or hateful comments or images posted online. Sample of 2,000 people. Source: The Alan Turing Institute (Vidgen, B., Margetts, H., Harris, A.), 2019. How much online abuse is there? [accessed 28 September 2023].
[597] The Economist Intelligence Unit. 2021, 'Measuring the Prevalence of Online gender-based violence against Women', [accessed 29 July 2024].
[598] Plan International, 2020, State of the World's Girls 2020: Free to be Online? [accessed 29 July 2024].

especially among certain groups such as women in the public eye, or women in the online gaming community. **Gender** also intersects with **age** and **race** as a risk factor, with evidence suggesting that young women, and those in **minority ethnic groups**, are at highest risk of harassment and abuse.

Functionalities and recommender system risk factors:

Several functionalities of User-to-user (U2U) services can be used in specific ways to perpetuate harassment, stalking and violent threats. While anonymity can be a source of protection, evidence suggests that **anonymous user profiles** may encourage harmful contact by making users feel freer to violate social norms. **User profiles**, and the information that is often displayed on them, can help facilitate stalking. Perpetrators can also gain unauthorised access to victims and survivors' accounts to impersonate them through their user profile.

Cases of harassment and stalking often involve perpetrators creating multiple and often **fake user profiles** to contact individuals against their will and to be omnipresent in their lives. Perpetrators can circumvent blocking and moderation by creating new accounts and their associated user profiles, thereby continuing to harass, stalk or threaten victims and survivors, causing significant fear and distress.

In some cases, perpetrators can leverage the **user connections** functionality by connecting with second and third-degree connections of the victim or survivor, to access content that is otherwise not publicly available, thereby giving a perpetrator visibility of a target's profile without connecting with them directly. User connections also enable perpetrators to build online networks through **network recommender systems**, which can be leveraged to facilitate harassment and abuse. Individual perpetrators can incite their network to join the abuse of an individual.[599] Additionally, **content recommender systems** can play a role in spreading harassment, threats and abusive content and in increasing the risks of perpetrators engaging in this content.

The **ability to post or send location information** can provide information that allows perpetrators to target and monitor victims and survivors for the purposes of harassment, stalking and threats of violence. **Reposting or forwarding content** can enable content to be circulated that is likely to provoke harassment or abuse from certain audiences. Abusive messages can also be communicated via **direct messaging,** in both private and public contexts. Abuse and harassment can also occur via **comments on content**.

Perpetrators also exploit **user tagging** to harass their victims and survivors by incessantly tagging their usernames in abusive and threatening messages.

[599] This is also known as a 'pile-on', in which an individual is attacked by a large number of users. This form of harassment can cause significant harm to victims and survivors.

Introduction

4.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the harassment, abuse, threats, stalking, and the threatening communications offences listed under 'Relevant offences' below; and
- the use of these services for the commission and/or facilitation of these offences (collectively, the 'risks of harm').

4.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the effect of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

4.3 The offences of harassment, stalking, threats and abuse relate to unwanted behaviours that can cause alarm and distress to other individuals or put them in fear of violence.[600] A case can involve several types of behaviour, hence the grouping of these offences for this chapter.

4.4 In this chapter:

- We discuss specific types of threats (such as threats to kill, rape threats, threats of violence) together as 'violent threats'[601]
- We use the term 'cyberstalking' in reference to literature that uses this term. While there is no legal definition of 'cyberstalking' and no specific legislation to address the behaviour, it is commonly used to refer to harassment and stalking taking place through electronic means such as the internet
- We refer to a range of sources of evidence, which are likely to interpret 'harassment' differently, and more broadly, than the specific priority offences being considered in this chapter.

4.5 Stalking, harassment, threats and abuse offences also occur as part of, or in conjunction with, several other online harms explored in the Register of Risks. Primary overlaps include Terrorism, Hate, Controlling or coercive behaviour, Intimate image abuse, Cyberflashing and Foreign interference.

Relevant offences

4.6 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding harassment, stalking, threats and abuse, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

[600] Crown Prosecution Service, 2023. Stalking or Harassment. [accessed 28 September 2023].
[601] The offences of fear or provocation to violence and threats to kill will be discussed together. This is because evidence of these threats perpetrated on services often relates to general threats to violence.

4.7 In this chapter, we consider the following offences:

- Making a threat to kill[602]
- Behaving in a threatening or abusive manner likely to cause fear or alarm[603]
- Using threatening, abusive or insulting words or behaviour, with intent to cause fear or provocation of immediate violence[604]
- Using threatening, abusive or insulting words or behaviour, or disorderly behaviour, or displaying any writing, sign or other visible representation which is threatening, abusive or insulting, with intent to cause a person harassment, alarm or distress[605]
- Using threatening, abusive or insulting words or behaviour, or disorderly behaviour, or displaying any writing, sign or other visible representation which is threatening or abusive, within the hearing or sight of a person likely to be caused harassment, alarm or distress nearby[606]
- Pursuing a course of conduct which amounts to harassment[607]
- Pursuing a course of conduct which amounts to stalking[608]
- Pursuing a course of conduct which puts a person in fear of violence[609]
- Pursuing a course of conduct amounting to stalking which puts a person in fear of violence, or causes serious alarm or distress[610]

4.8 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of the offences listed above (and in relation to offences in Scotland, being involved art and part in the commission of those offences).

4.9 The priority offences listed above cover a range of behaviours:

- **Threats** and **threatening behaviour, abuse and abusive behaviour** – this includes threats to kill a person, other violent threats and threats intended to cause fear of, or provoke, immediate violence, and general abusive behaviour that may cause a person harassment, alarm or distress.
- **Harassment** is when a person engages in a course of conduct (a minimum of two instances), which amounts to harassment of another person. Harassment includes causing that other person alarm or distress. The course of conduct may involve the same or different behaviours on each occasion. On its own, an instance may amount to another offence such as hate or abuse (see Hate chapter for further information).

---

[602] Section 16 of the Offences against the Person Act 1861.
[603] Section 38 of the Criminal Justice and Licensing (Scotland) Act 2010 (asp 13).
[604] Section 4 of the Public Order Act 1986.
[605] Section 4A of the Public Order Act 1986.
[606] Section 5 of the Public Order Act 1986.
[607] Section 2 of the Protection from Harassment Act 1997; Article 4 of the Protection from Harassment (Northern Ireland) Order 1997 (S.I. 1997/1180 (N.I. 9)).
[608] Section 2A of the Protection from Harassment Act 1997; Section 39 of the Criminal Justice and Licensing (Scotland) Act 2010 (asp 13).
[609] Section 4 of the Protection from Harassment Act 1997; Article 6 of the Protection from Harassment (Northern Ireland) Order 1997 (S.I. 1997/1180 (N.I. 9)).
[610] Section 4A of the Protection from Harassment Act 1997.

Harassment can also be by two or more people against an individual, or harassment against more than one victim or a group.[611]

- **Stalking** similarly involves a course of conduct and is a specific form of harassment. It may be understood as a pattern of fixated, obsessive, unwanted and repeated (FOUR) behaviour which is intrusive.[612] Examples of behaviours associated with stalking include following a person or contacting or attempting to contact a person.

4.10 Threatening and abusive behaviour can occur on individual occasions when a perpetrator communicates with an individual in a way that causes alarm or distress. Repeated threatening or abusive behaviours can amount to stalking or harassment offences. These occur when a perpetrator, or multiple perpetrators, undertake a course of action which can build up fear and distress in the target individual. This course of action often comprises both online and offline behaviours.[613] Harassment causing fear of violence, stalking causing fear of violence, and stalking causing serious alarm or distress and having a substantial adverse impact on the victim's day-to-day activities are more serious forms of harassment and stalking.

4.11 Stalking and harassment online can differ from offline contexts, relying on specific technological affordances and dynamics. For example, stalking someone online requires no physical presence, and it can be easier to remain anonymous, and to incite others to commit harassment in place of the main perpetrator. The online environment can be used to locate personal information about a person, to communicate with them, or as a means of surveillance of the person.[614]

4.12 Illegal content online may manifest in any number of ways, including text, audiovisual content and images; this content may be a one-off or part of a pattern of content (for example, sending abusive or threatening messages, images or videos on U2U services).

4.13 For more details on how services can assess whether content amounts to priority illegal content and more general illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

How harassment, abuse, threats and stalking offences manifest online

4.14 This section is an overview which looks at how harassment, abuse, threats, stalking, and the threatening communications offences manifest online, and how individuals may be at risk of harm.

4.15 The evidence suggests that understanding how these harms manifest online requires considering how they often occur together. Analysis by the Crown Prosecution Service of stalking prosecutions has shown that most offences are committed by abusive ex-

[611] Crown Prosecution Service, 2023. Stalking or Harassment. [accessed 28 September 2023].
[612] Crown Prosecution Service, 2023.
[613] A survey conducted by the Suzy Lamplugh Trust, a UK-based charity specialising in stalking, found that 75% of respondents who had experienced stalking had experienced both online and offline stalking behaviours: Suzy Lamplugh Trust. 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].
[614] Crown Prosecution Service, 2023.

partners[615], while other data suggests that a significant proportion of intimate partner stalking victims are stalked during the relationship.[616] [617] Research by the Victims' Commissioner for England and Wales found that victims of online abuse often experience multiple types of other harms – with victims of cyberstalking in particular experiencing more harms on average.[618]

4.16 Additionally, while the chapter primarily considers how these harms manifest online, the evidence suggests that the online manifestation of these harms is often linked to their incidence offline. The National Stalking Helpline reported that 100% of the stalking cases reported to the helpline now involve a 'cyber' element, and a survey conducted by the Suzy Lamplugh Trust found that 3 out of 4 (75%) respondents who had experienced stalking, experienced both online and offline stalking behaviours.[619]

Harassment, abuse and threats

4.17 Crime surveys provide an indication of the number of people affected by online harassment. The Scottish Crime and Justice Survey asks respondents about experiences of being insulted or harassed online. For 2017 to 2018 it reported that 14% of adults reported experiencing harassing or intimidating behaviour and that, of these, 16% had encountered such behaviour digitally.[620] However, it is worth noting that the UK national institute for data science and artificial intelligence, the Alan Turing Institute, describes the evidence for understanding the presence of online abuse more generally as *"fragmented, incomplete and inadequate*."[621]

4.18 Estimates of the presence of online abuse more broadly are limited, and interpersonal online abuse, as distinct from abuse towards protected groups (explored in the Hate chapter), is rarely measured. However, the available evidence suggests that a substantial portion of the UK population is exposed to online abuse in some capacity. In Ofcom's research tracking people's experiences online, 12% of people aged 13+ surveyed, reported they had witnessed 'one off abusive behaviour or threats' in the past four weeks.[622] A Pew Research study in 2020 found that 41% of Americans had reported personally experiencing

[615] Crown Prosecution Service, 2020. Stalking analysis reveals domestic abuse link. [accessed 10 October 2024].
[616] Paragon, 2024. Domestic abuse and stalking. [accessed 10 October 2024].
[617] The Domestic Abuse Commissioner also notes that stalkers in a domestic context (intimate partners or ex-intimate partners) ought to be recognised as a distinctive category in respect of the prevalence and risk posed to victims. Source: Domestic Abuse Commissioner's response to November 2023 Illegal Harms Consultation.
[618] The Victims' Commissioner (Storry, M., Poppleton, S.), 2022. The Impact of Online Abuse: Hearing the Victims' Voice. [accessed 25 July 2024].
[619] Suzy Lamplugh Trust. 2021. Unmasking stalking: a changing landscape. [accessed 25 July 2024].
[620] Scottish Government. 2018. Scottish Crime and Justice Survey 2017/18. [accessed 28 September 2023].
[621] The Alan Turing Institute (Vidgen, B., Margetts, H., Harris, A.), 2019. How much online abuse is there? [accessed 28 September 2023].
[622] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].

online abuse and harassment.[623] The Oxford Internet Survey in 2019 reported that 27% of people had been exposed to abusive content (including hate speech).[624]

4.19 Online abuse and harassment – particularly that which is experienced by women and girls – is a global issue. In a study surveying 4500 women globally, 38% of women aged 18 to 74 stated that they had personal experiences of "online violence" and 65% reported knowing other women (in their personal and professional networks) who had been targeted online.[625] In Plan International's study of over 14,000 girls and women aged 15 to 25, 58% of respondents reported they experienced some form of online harassment.[626]

4.20 Evidence exclusively measuring the presence of violent threats is limited. Violent threats are listed as examples of abusive behaviour, measured in the round. Pew Research data from the US showed a doubling of physical threats on US adults between 2014 and 2020: from 7% to 14%.[627]

4.21 An important aspect of abuse and harassment online – as opposed to offline – is the potential for these harms to be committed in ways that are more, or more easily, coordinated, done at large-scale, and long-lasting. For instance, U2U services that have hashtag functionalities can facilitate networked harassment through enabling harassers to rapidly enlist others, quickly spread inciteful content, and easily coordinate targeting campaigns.[628] Research suggests that the public nature of such campaigns can even result in wider collective harms that can be especially detrimental to those who share the identity of the targeted groups even when they are not direct targets of the harassment.[629]

Stalking

4.22 The 2022 Crime Survey for England and Wales found that 16.6% of people over 16 had experienced stalking, with 5.7% having experienced stalking with an online element.[630] The National Policing Statement for Violence Against Women and Girls (VAWG) found in 2024 that stalking and harassment accounts for 85% of all online and tech-enabled offences.[631]

---

[623] 41% of respondents reported experiencing at least one of physical threats, stalking, sexual harassment, sustained harassment, purposeful embarrassment, or offensive name-calling. Note: Where possible, UK data has been used throughout this chapter. However, when this is limited, evidence for comparable cultures has been used, namely the US, Australia and Canada. Where evidence is not UK-based, this will be clearly stated. This US-based study from Pew Research includes single occasions as measures of harassment, although in the UK the harassment offence is a course of conduct occurring on two or more occasions. Source: Pew Research (Vogels, E.), 2021. The State of Online Harassment. [accessed 28 September 2023].

[624] 27% of respondents had seen cruel or hateful comments or images posted online. Sample of 2,000 people. Source: The Alan Turing Institute (Vidgen, B., Margetts, H., Harris, A.), 2019. How much online abuse is there? [accessed 28 September 2023].

[625] The Economist Intelligence Unit. 2021, 'Measuring the Prevalence of Online gender-based violence against Women', [accessed 29 July 2024].

[626] Plan International, 2020, State of the World's Girls 2020: Free to be Online? [accessed 29 July 2024].

[627] Pew Research (Vogels, E.), 2021. The State of Online Harassment. [accessed 28 September 2023].

[628] Gamergate, a networked harassment campaign that targeted women in the video game industry, is a high-profile example of this that is often cited when studying these issues. Source: Romano, A., 2021. What we still haven't learned from Gamergate, Vox, 7 January. [accessed 29 July 2024].

[629] Fox J., Cruz C., & Lee J., 2015. Perpetuating online sexism offline: Anonymity, interactivity, and the effects of sexist hashtags on social media. *Computers in Human Behavior*, 52, 436-442. [accessed 19 November 2024].

[630] Office for National Statistics, 2022. Stalking: findings from the Crime Survey for England and Wales. [accessed 28 September 2023].

[631] National Police Chief Council, 2024. Violence Against Women and Girls (VAWG) National Policing Statement. [accessed 29 July 2024].

Research shows that stalking rarely occurs online only. Studies indicate that stalking which starts online tends to move offline and stalking that occurs predominantly offline will also have online elements. The National Stalking Helpline reports that 100% of the stalking cases reported to the helpline now involve a 'cyber' element, and a study into modes of cyberstalking and cyber harassment found that 'proximal' stalking which begins offline but also includes online elements was the most common among 278 victims in the UK. Information gained through online stalking (such as location coordinates from default geo-tagging of images shared on social media) can enable offline stalking and violence.[632] [633] [634]

4.23 The COVID-19 pandemic is found to have heightened the risk of stalking perpetrated using online services, and specifically on social media services.[635] Perpetration on U2U services has also increased; before the pandemic, text messages and direct messages were the most common mode of stalking. After the pandemic, text message and direct messages remained stable, but the proportion of those experiencing stalking on social media services increased.[636]

Risks of harm to individuals presented by these offences

4.24 Harassment, threats, stalking and abuse offences manifest as a wide variety of online behaviours that cause fear, distress and alarm to victims and survivors.

4.25 Some abusive behaviours are more visible, such as sending online verbal abuse or posting content intended to publicly humiliate individuals. Perpetrators also cause distress in more ongoing and covert ways, such as by monitoring or controlling targets' accounts, impersonating them, or inciting others to participate in the abuse.[637] Stalking and harassment cases can involve a repeated behaviour, such as persistent unwanted messages on social media services, or a range of different behaviours, such as sending abusive messages as well as monitoring victims and survivors' accounts.

4.26 Identifying content that causes fear or distress demands an understanding of the context. For example, sending a picture of someone's front door or workplace address might seem innocuous, but may be highly threatening, by making victims and survivors aware that the perpetrators can access them physically.[638] Similarly, hyper-local knowledge of context, culture and nuance (such as coded references to individuals or events) can be necessary in identifying posts that are perceived as direct threats to violence by gang-affiliated individuals.[639]

---

[632] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 25 July 2024].
[633] Brown A., Gibson M., & Short E. 2017. Modes of Cyberstalking and Cyberharassment: Measuring the negative effects in the lives of victims in the UK *Annual Review of Cybertherapy and Telemedicine.* [accessed 19 November 2024].
[634] Sheridan L., Blaauw E., &Davies G. 2003. Stalking: Knowns and Unknowns. *Trauma, Violence & Abuse* 4(2). [accessed 19 November 2024].
[635] Titherade, N. and Thomas, E., 2021. Stalking rises during Covid pandemic - police, BBC News, 13 July. [accessed 12 July 2023].
[636] Suzy Lamplugh Trust, 2021.
[637] Crown Prosecution Service, 2023. Stalking or Harassment. [accessed 28 September 2023].
[638] Refuge, 2021. Unsocial Spaces. [accessed 28 September 2023].
[639] Patton, D. U., Pryooz, D., Decker, S., Frey, W. R. and Leonard, P., 2019. When Twitter Fingers Turn to Trigger Fingers: a Qualitative Study of Social Media-Related Gang Violence. [accessed 28 September 2023]; Crest, 2022. Calouri, J., Hutt, O., Olajide, P. and Kirk, E, 2022;

4.27 Anyone can be affected by these behaviours, but evidence suggests that women are disproportionately affected.[640] Alongside gender, other demographic factors such as race, ethnicity, and faith increase the risks of harm and are explored in greater detail in the Risk factors: user base section below.

4.28 The risk from aggravated forms of harassment and stalking, and violent threats offences (which could include threats to kill, rape or perform acts of violence) is particularly acute for survivors of sexual and domestic abuse. Where perpetrators of these offenses make threats, evidence suggests that 20 to 50% will follow that threat through.[641] The Register of Risk also refers to this offence in the Controlling or coercive behaviour and Sexual exploitation of adults chapters.

Harassment, abuse and threats

4.29 Harms caused by these offences are often severe and varied. A report from the Victims Commissioner found that online abuse and harassment can cause significant harm to victims, including feelings of anxiety, depression, and social withdrawal.[642] While another study found that individuals who had been exposed to threatening and/or obscene messages in the past year were 5.49 times more likely to attempt suicide compared to the rest of the population.[643]

4.30 Experiencing abuse and harassment can have a silencing effect, making victims and survivors feel unsafe in expressing themselves on social media services. Human rights organisation Amnesty International found that 24% of women experiencing online abuse and harassment said they stopped posting their opinions on certain issues.[644] A study from 2016 found that 27% of US internet users censor their own online posts for fear of being harassed.[645] This silencing effect can harm victims and survivors' careers. A study of women journalists found that those facing abuse and harassment reported making themselves less visible (38%), missing work (11%), leaving their jobs (4%), with some deciding to abandon journalism altogether (2%).[646]

4.31 A common consequence of online harassment for victims and survivors is isolation or disconnection from their communities, whether because of the strain the harassment has put on their close relationships, or because their harassment has made them feel more cut off from avenues for communication and information-seeking. A study from the US in 2016

---

[640] The Crime Survey for England and Wales found that while almost one in ten men (9.5%) have experienced stalking, almost a quarter of women (23.3%) over the age of 16 have experienced this offence. The Crime Survey for England and Wales also collects information on the prevalence of stalking with an online element, and found that 8.3% of women had experienced this, and 3.1% of men. Source: Office for National Statistics, 2022. Stalking: findings from the Crime Survey for England and Wales. [accessed 28 September 2023].

[641] McEwan, T. E., Mullen, P. E., MacKenzie, R. D., Ogloff, J. R. P. 2009. Violence in stalking situations. *Psychol Med*. [accessed 13 September 2024].

[642] Victims Commissioner for England and Wales. The Impact of Online Abuse: Hearing the Victims' Voice. [accessed 24 October 2024].

[643] This is when adjusted for socioeconomic status. Source: McManus, S., Bebbington, P. E., Tanczer, L., Scott, S. and Howard, L. M. 2021. Receiving threatening or obscene messages from a partner and mental health, self-harm, and suicidality: results from the Adult Psychiatric Morbidity Survey. *Social Psychiatry and Psychiatric Epidemiology: the international journal for research in social and genetic epidemiology and mental health services.* [accessed 29 July 2024]

[644] Amnesty International, 2017. Social media can be a dangerous place for UK women. [accessed 28 September 2023].

[645] Data & Society Research Institute/CiPHR (Lenhart, A., Ybarra, M., Zickuhr, K. and Price-Feeney, M.), 2016. Online Harassment, digital abuse and cyberstalking in America. [accessed 28 September 2023].

[646] UNESCO (Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S.), 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].

found that 40% of victims and survivors of online harassment said they had experienced at least one of these types of isolation or disconnectedness as a result:[647]

- 27% of victims and survivors experienced trouble in a relationship or friendship because of something that was posted about them online;
- 20% had shut down an online account or profile because of online harassment or abuse; and
- 13% of victims and survivors felt less connected to information and 13% felt less connected to friends or family because their phone or internet use was limited because of harassment or abuse.

4.32 The impact on individuals is highly dependent on the personal characteristics of victims and survivors and the circumstances in which abuse and harassment occur. Evidence relating to a variety of personal characteristics is explored in the 'user base demographics' section below, but it is important to recognise the particular role that gender plays. Amnesty international found that more than half of UK women (55%) who experienced abuse or harassment online experienced stress, anxiety or panic attacks following the abuse, while 36% of women said it made them feel that their physical safety was threatened.[648] Research from the USA found that of those who identified as having experienced harassment or abuse, women were almost three times as likely as men to say the harassment made them feel scared, and twice as likely to say the harassment made them feel worried.[649] Fourteen percent of men found their most recent experience of online harassment 'very' or 'extremely' upsetting, compared to 34% of women.[650] Impacts can also vary for victims and survivors of different ethnicities.[651]

Stalking

4.33 Stalking can be a contributing factor for offline violence up to and including murder. One study found stalking present in 94% of homicide cases where the victim was a woman specifically targeted by the perpetrator.[652] Another study of intimate partner homicides found that in 85% of attempted homicides and 76% of completed homicides, the victims were stalked.[653] The evidence indicates that most cases of stalking will feature a perpetrator relationship to the victim, although some perpetrators will stalk a victim they have not had any direct relationship with. A survey by the Suzy Lamplugh Trust found that

---

[647] Data & Society Research Institute/CiPHR (Lenhart, A., Ybarra, M., Zickuhr, K. and Price-Feeney, M.), 2016. Online Harassment, digital abuse and cyberstalking in America. [accessed 28 September 2023].
[648] Amnesty International, 2017. Social media can be a dangerous place for UK women. [accessed 28 September 2023].
[649] Data & Society Research Institute/CiPHR, 2016.
[650] Pew Research, 2021. The State of Online Harassment. [accessed 28 September 2023].
[651] A study in Australia explored how ethnicity and gender intersect, with specific impacts (such as threats of deportation, 'honour' killing, or culturally-specific humiliation) affecting women with ethnically diverse backgrounds. Source: eSafety Commission Australia, 2019. eSafety for Women from Culturally and Linguistically Diverse Backgrounds. [accessed 28 September 2023].
[652] The authors reviewed 358 cases of homicide where the victim was a woman, and collected information to identify key stalking, control and risk markers in each case. They excluded cases where the victim was not specifically targeted, i.e. cases of mistaken identity, where the homicide outside the UK and circumstances could not be verified, or where the homicide occurred in the course of another crime such as a robbery. Source: Suzy Lumplugh Trust (Monckton Smith, J., Szymanska, K., and Haile, S.), 2017. Exploring the Relationship between Stalking and Homicide. [accessed 19 November 2024].
[653] McFarlane, J., Campbell, J.C., Wilt, S., Ulrich, Y., & Xu, X. 1999. Stalking and Intimate Partner Femicide. Homicide Studies, 3(4), 300-316. [accessed 19 November 2024].

61% of stalking victims are being or have been stalked by a former partner, with only 7% stating that the perpetrator was a stranger.[654] The National Stalking Helpline further reports that *"45% of those who contact the Helpline are being stalked by a former partner and further third have had some sort of prior acquaintance with their stalker."*[655]

4.34 A systematic review of the literature on the experiences of victims and survivors found that cyberstalking can have adverse functional, physiological, and psychological effects:[656]

- Functional – including lower professional or academic performance and financial costs.
- Physiological – including changes to eating habits and disrupted sleep patterns.657
- Psychological – including mental and emotional distress, fear, depression, anxiety, withdrawal from social activities.

4.35 A 2020 survey found that 94% of stalking victims stated that the stalking they experienced negatively impacted their mental health, and another survey studying the impact of cyberstalking found that 50% of victims developed post-traumatic stress disorder (PTSD).[658] [659] Research describes how stalking victims experience hypervigilance – a state of heightened alertness and anxiety – whereby they feel as if perpetrators "are able to invade every aspect of their live, at any time of day or night, and across locations."[660]

Evidence of risk factors on user-to-user services

4.36 We consider that the risk factors below are likely to increase the risks of harm relating to harassment, stalking, threats and abuse offences.

Risk factors: Service types

4.37 Research indicates that the following types of services are particularly relevant to the offences of harassment, stalking threats and abuse: social media services, online gaming services, online dating services and private messaging services.

Social media services

4.38 Social media services are a risk factor for these offences. Stalking, harassment and violent threats can occur on various online services such as online dating services and gaming

---

[654] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 29 July 2024].
[655] Suzy Lamplugh Trust, n.d. What is stalking? | Suzy Lamplugh Trust. [accessed 29 July 2024].
[656] Kaur, P., Dhir, A., Tandon, A., Alzeiby, E.A. and Abohassan, A.A., 2020. A systematic literature review on cyberstalking. An analysis of past achievements and future promises, *Technological Forecasting and Social Change,* 163. [accessed 28 September 2023].
[657] These impacts are described by a feminist writer and campaigner who faced an extensive online abuse and harassment: *"At its height I struggled to eat, to sleep, to work. I lost about half a stone in a matter of days. I was exhausted and weighed down by carrying these vivid images, this tidal wave of hate around with me wherever I went... the psychological fall-out is still unravelling*" Source: Criado Perez, C. 2015. Caroline Criado-Perez's speech on cyber-harassment at the Women's Aid conference. *The New Statesmen*, 27 September. [accessed 28 September 2023].
[658] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 29 July 2024].
[659] Short E. and Maple C., 2011. The impact of cyberstalking: review and analysis of the ECHO pilot project. *Proceedings of the IADIS International Conferences – Web Based Communities and Social Media.* [accessed 29 July 2024].
[660] Flynn A., Powell A., Hindes S. (2021). Technology-facilitated abuse: A survey of support services stakeholders. [accessed 29 July 2024].

services, but it happens a lot more on social media services.[661] According to UNESCO, "*social media companies are the main enablers of online violence against women journalists*", which includes sexual violence and 'gendered profanities'.[662]

Online gaming services

4.39 Several sources of evidence suggest that online gaming services are a risk factor. A survey on American gamers found that the harassment experienced by adult gamers is both 'alarmingly high' and on the rise. Five out of six adults (83%) aged 18 to 45 had experienced harassment in online multiplayer games, while 71% had experienced severe abuse, including physical threats, stalking, and sustained harassment.[663] Online gaming services or 'networked gaming' has also been identified as sites of 'normalised harassment', where name-calling or insults are part of the culture.[664] The risk can be elevated by livestreaming; gamers in a 2021 study believed that it leads to more bullying behaviour.[665]

4.40 Harassment while gaming is a common experience for both women and men. However, a study exploring female gamers' experiences of social support while playing online games found that "*female gamers often report experiencing harassment whilst playing online*."[666] Strategies used by women to deal with online game-related harassment include leaving online gaming, avoiding playing with strangers, or camouflaging their gender.[667] There is less evidence on male or intersectional experiences, though research suggests that Black people experience more harassment and threats on gaming services and that gaming services can host communities which promote racial and anti-Black violence. [668]

Online dating services and private messaging services

4.41 Online dating services and private messaging services may also be risk factors. Negative interactions with other users are common on dating sites or apps. A study in the US found that 35% of American online dating users report being sent a sexually explicit message or image they did not ask for, 28% report being called an offensive name, and 9% report receiving threats to physically harm them. These figures are significantly higher for young

---

[661] Some research shows that stalking through social media services is becoming more common. A study by the Suzy Lamplugh trust found that stalking using social media jumped from 59% before the first Covid-19 pandemic lockdown to 82% after it. Note small sample of 111 victims of stalking in survey. Source: Suzy Lamplugh Trust. 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].
The NPCC reports that the majority of online-facilitated VAWG incidents that are reported to the police is classified as stalking, and is committed via social media services. See 'How harms manifest online' section for more information. Source: National Police Chiefs Council, 2023. Violence Against Women and Girls: Strategic Threat Risk Assessment 2023. [accessed 28 September 2023].

[662] Posetti, J., Shabbir, N., Maynard, D., Bontcheva, K. and Aboulez, N., 2021. The chilling effect. Global trends in online violence against women. [accessed 28 September 2023].

[663] Anti-defamation League, 2021. Hate is No Game: Harassment and Positive Social Experiences in Online Games 2021. [accessed 28 September 2023].

[664] Marwick, A, 2021. Morally Motivated Networked Harassment as Normative Reinforcement. [accessed 28 September 2023].

[665] See 'Risk factors: functionalities and recommender systems' section for more information. Source: McLean, L. and Griffiths, M. D., 2018. Female Gamers' Experience of Online Harassment and Social Support in Online Gaming: A Qualitative Study, *International Journal of Mental Health and Addiction*, 17, 970-994. [accessed 28 September 2023].

[666] McLean, L. and Griffiths, M. D., 2018. Female Gamers' Experience of Online Harassment and Social Support in Online Gaming: A Qualitative Study, *International Journal of Mental Health and Addiction*, 17, 970-994. [accessed 28 September 2023].

[667] See 'Risk factors: user base' section for more information. Source: Cote, C, 2017. ''I Can Defend Myself'': Women's Strategies for Coping With Harassment While Gaming Online, *Games and Culture*, 12(2). [accessed 28 September 2023].

[668] Gray, K.L. 2021. Intersectional Tech: black users in digital gaming. *LSU Press.*

women (aged 18 to34), with 57% receiving an explicit message, 44% being called an offensive name, and 19% receiving threats to physical harm.[669]

4.42 Our evidence suggests that the harassment of public figures and stalking often occurs through direct messaging, which is a core functionality of private messaging services. One study showed that texts or direct messages were the most common digital stalking behaviour,[670] and another that 48% of female journalists had been harassed by unwanted private messages.[671]

Risk factors: user base

User base demographics

4.43 The following section outlines key evidence of user base demographic factors and risks of harm, which can include protected characteristics (see glossary for definition). Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

4.44 The data suggests that userbase characteristics: **age, gender, race and ethnicity** could affect the risks of harm to individuals.

4.45 The evidence suggests that young adults are more vulnerable to stalking, harassment and threats of violence. Ofcom's Online Experiences Tracker showed that users aged 18 to 34 were more likely than the average internet user to have seen or experienced stalking cyberstalking or harassment in the past four weeks (3% versus 2%).[672] Services with many younger adult users may therefore be disproportionately open to the risks of harm from harassment, stalking, threats and abuse.

4.46 Pew Research data from the US shows starker differences between age groups in both the prevalence and the severity of harassment, with younger users experiencing more harassment, specially forms of harassment that the study classifies as 'severe'.[673]

4.47 Gender is a significant risk factor for these offences, with evidence suggesting that both prevalence and risk of harm to individuals are higher among women.

4.48 According to a poll conducted for Amnesty International in June 2017, one in five women in the UK have suffered online abuse or harassment, increasing to one in three for young women aged 18 to 24. More than a quarter (27%) of women of any age experiencing abuse or harassment received some form of threat (direct or indirect) of physical or sexual

---

[669] Online dating users refers to respondents who say they have ever used an online dating site or app (n=2,094). Source: Anderson, M., Vogels, E., Turner, E, 2020. The Virtues and Downsides of Online Dating. [accessed 28 September 2023].
[670] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].
[671] See 'Risk factors: functionalities and recommender systems' section for more information. Source: UNESCO (Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S.), 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].
[672] Comprises waves 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of using online services. [accessed 28 September 2023].
[673] According to this study, 64% of US adults aged 18-30 had experienced any form of harassment, compared to 41% for the whole adult population. Nearly half (48%) of 18-30 year olds had experienced behaviours classified by the study as more severe. These include being include being physically threatened, stalked, sexually harassed or harassed for a sustained period of time. Source: Pew Research, 2021. The State of Online Harassment. [accessed 28 September 2023].

assault.[674] A study of misogynoir[675] on several social media platforms found that 20% of the one million posts collected about women were highly toxic.[676]

4.49 Comparative statistics between genders are limited. US data from 2017 suggested that the prevalence of harassment was similar between men and women, but that women were more severely affected: women who had experienced harassment were more than twice as likely to say the most recent incident was very or extremely upsetting. High prevalence among women is linked to online misogyny. Amnesty found that nearly half of UK women who experienced online abuse or harassment received sexist or misogynistic comments (47%).[677]

4.50 Evidence suggests that women in public-facing professions are particularly affected by harassment, threats and abuse, and that Black and Asian women in such professions are at higher risk of experiencing these harms. A UNESCO study found that 25% of the women journalists sampled had received threats of physical violence and 18%, threats of sexual violence.[678] A study by the Inter-Parliamentary Union of women parliamentarians, profiled across 36 countries, showed that they had been harassed (defined as insistent and uninvited behaviour) and had received threats of rape, beatings or abduction, mostly through email or social media services.[679] Research by Amnesty International also shows that Black, Asian, and Minority Ethnic (BAME) women MPs receive significantly more abusive messages than their white counterparts.[680]

4.51 The National Crime survey found that 8.3% of women and 3.1% of men reported having experienced cyberstalking specifically.[681] The Suzy Lamplugh Trust similarly found that 79% of the victims and survivors supported by the National Stalking Helpline in the past year identified as female, indicating that female-identifying victims and survivors are disproportionately seeking help.[682]

4.52 Comparative analysis into abuse and harassment in gaming is limited. However, studies report on the prevalence of misogyny in these spaces. Experiences of harassment are a common theme in discussion groups for female gamers. Some describe the experience of being stalked both online and offline.[683] Strategies to deal with negative experiences include leaving online gaming, avoiding playing with strangers, or disguising their gender,

---

[674] Amnesty International, 2017. Social media can be a dangerous place for UK women. [accessed 28 September 2023].
[675] "Misogynoir" is a term used to describe discrimination against Black women.
[676] Glitch, 2023. The Digital Misogynoir Report: Ending the dehumanising of Black women on social media. [accessed 29 July 2024]
[677] Amnesty International, 2017. Social media can be a dangerous place for UK women. [accessed 28 September 2023].
[678] UNESCO (Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S.), 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].
[679] Inter-Parliamentary Union, 2016. Sexism, harassment and violence against women parliamentarians. [accessed 28 September 2023].
[680] Amnesty International UK. 2017. Black and Asian Women MPs Abused More Online. [accessed 29 July 2024].
[681] Office for National Statistics, 2022. Stalking: findings from the Crime Survey for England and Wales. [accessed 28 September 2023].
[682] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].
[683] See 'Risk factors: functionalities and recommender systems' section for more information. Source: McLean, L. and Griffiths, M. D., 2018. Female Gamers' Experience of Online Harassment and Social Support in Online Gaming: A Qualitative Study, *International Journal of Mental Health and Addiction*, 17, 970-994. [accessed 28 September 2023].

suggesting that the risks of harm to individuals are higher for players who present as female in these environments.[684]

4.53 Age and gender intersect as risk factors, with young women particularly vulnerable to online stalking and harassment. The National Crime survey finds stalking to be most prevalent among women aged 25 to 34 (5.1%), and cyberstalking among women aged 20 to 24 (3.5%).[685] US data from 2020 finds that 33% of women under 35 report being sexually harassed online, while 11% of men under 35 say the same.[686]

4.54 Users from minority ethnic groups are at higher risk of harm. Ofcom's Online Experiences Tracker (OET) data shows higher prevalence of stalking, cyberstalking, and harassing behaviours among ethnic minority groups, with 8% having seen or experienced 'stalking, cyberstalking or harassing behaviour' in the past four weeks, compared to 5% of white respondents.[687]

4.55 The risk of harm to individuals is higher due to the prevalence of racist sentiment in online spaces. According to Pew Research data, online harassment on grounds of race and ethnicity is increasing. In 2017, 19% of those who had been harassed online cited race and ethnicity as the reason they were targeted, rising to 29% in 2020.[688] Fifty-four percent of Black and 47% of Hispanic online harassment targets said they were harassed due to their race or ethnicity, compared with 17% of white targets.[689]

4.56 Race and gender are also intersecting risk factors. For example, research has demonstrated that non-White women are significantly more likely to be targets of abusive messages on online services[690] [691]; and that online misogynoir is prevalent on social media services.[692] In the National Crime survey for England and Wales, reports of stalking are also most prevalent among Black or Black British women (6.5%).[693]

4.57 Ofcom data suggests that this online harm is experienced more by internet users in certain religious groups, with claimed prevalence among Hindus at 12% compared to 4% for members of the Churches of England, Scotland or Ireland.[694]

---

[684] Cote, C, 2017. ''I Can Defend Myself'': Women's Strategies for Coping With Harassment While Gaming Online, *Games and Culture*, 12(2). [accessed 28 September 2023].
[685] Office for National Statistics, 2022. Stalking: findings from the Crime Survey for England and Wales. [accessed 28 September 2023].
[686] Pew Research, 2021. The State of Online Harassment. [accessed 28 September 2023].
[687] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].
[688] Pew Research, 2021.
[689] Pew Research, 2021.
[690] Black women are 84% more likely to be targets of abusive tweets than white women and 60% more likely to receive problematic tweets. Source: Glitch response to 2022 Ofcom Call for Evidence: First phase of online safety regulation; A US study collected a sample of Twitter data between 2015 and 2017. Analysing these 25,000 tweets, they found it took on average 18 seconds to detect an insulting, negative tweet directed at Black women, and 16 seconds to locate such a message aimed at Latina women, leading the authors to conclude that aggressive messages towards women of colour were easily accessible and visible on the social media platform. Source: Francisco, S. and Felmlee, D.H., 2022. What Did You Call Me? An Analysis of Online Harassment Towards Black and Latinx Women, *Race and Social Problems*, 14, 1-13. [accessed 28 September 2023].
[691] Glitch, 2023. The Digital Misogynoir Report: Ending the dehumanising of Black women on social media. [accessed 29 July 2024]
[692] Glitch, 2023.
[693] Office for National Statistics, 2022.
[694] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].

4.58 Stalking, cyberstalking or harassing behaviour is higher among those who identify as bisexual. Ofcom data found that while 5% of heterosexuals claimed experience of this harm, bisexuals were more likely to do so, at 11%.[695] Although not comparative, other evidence suggests that online harassment is higher among LGBTQ+ populations; higher still for certain groups within this community. Data from the EU Agency for Fundamental Rights suggests that 22% of LGBTQ+ populations in the UK have experienced cyber-harassment for any reason in the past five years, rising to 32% for trans people.[696] A further study on LGBTQ+ young people in Scotland found that a higher percentage of the lesbian and gay participants have experienced online bullying compared to their bisexual counterparts.[697] Additionally, OET data found that reports of stalking, cyberstalking or harassing behaviour are higher among transgender women and non-binary people (16%) compared to cisgender respondents (4%).[698]

4.59 Data from the USA compares the LGBTQ+ and heterosexual populations' experience of harassment and stalking, finding LGBTQ+ populations to be significantly more at risk. According to 2016 US research, 33% of the LGBTQ+ individuals sampled had been sexually harassed online, compared to 6% of heterosexual people. Thirty-one per cent reported being physically threatened (compared to 10%), and 31% reported being stalked online (compared to 7%).[699]

4.60 Ofcom data suggests that this online harm is experienced more by internet users with limiting conditions (8%), and those with mental health conditions (11%) compared to 3% with no such conditions.[700]

Risk factors: Functionalities and recommender systems

User identification

User profiles

4.61 The ability to create a user profile is a risk factor for these offences. Perpetrators can gain access to victims' and survivors' accounts, and then impersonate them through user profiles associated with those accounts.[701]

4.62 User profiles, and the information that is often displayed on them, can also be made available to networks of other users. Access to this content on a service can be used to commit or facilitate stalking and harassment offences. Without privacy settings in place, information such as an individual's preferences, activities and whereabouts can be open to the public. Monitoring this information is not an offence but can facilitate stalking and harassment. Evidence shows that victims and survivors are likely to have had their activities

[695] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].
[696] FRA EU Agency for Fundamental Rights, 2020. LGBTI Survey Data Explorer. [accessed 28 September 2023].
[697] LGBT Youth Scotland. Life in Scotland for LGBT Young People in 2022. [accessed 24 October 2024].
[698] Note this data set comprised Wave 1 and 2 of the Online Experiences Tracker in order to provide a large enough sample size for this analysis. Source: Ofcom, 2023. Experiences of using online services. [accessed 28 September 2023].
[699] Data & Society Research Institute/CiPHR (Lenhart, A., Ybarra, M., Zickuhr, K. and Price-Feeney, M.), 2016. Online Harassment, digital abuse and cyberstalking in America. [accessed 28 September 2023].
[700] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].
[701] Clevenger, S. and Gilliam, M, 2020. Intimate partner violence and the internet: Perspectives. Chapter in Holt, T. J. and Bossler, A. M. (eds.), The Palgrave Handbook of International Cybercrime and Cyberdeviance. [accessed 28 September 2023].

monitored on social media services; the US Department of Justice in 2019 found this to be true of 31.9% of stalking victims and survivors.[702]

Fake user profiles

4.63 Perpetrators can create multiple, often fake user profiles to help create the impression of omnipresence in victims' and survivors' lives, causing the alarm or distress that defines these offences.[703] These profiles are often used to harass victims and survivors. If the account is blocked, new accounts and their associated user profiles can be created that often impersonate other individuals.

4.64 Several studies report perpetrators of abuse and harassment creating fake user profiles.[704] For example, a study by domestic abuse charity Refuge reports one individual attempting to block her former partner, only to find over 120 fake accounts created by him over a few weeks to continue harassing her.[705] In a high-profile UK case, a man was convicted of continually harassing women online through creating false profiles, harassing 62 women over a ten-year period.[706] A feminist writer and campaigner described her sense of powerlessness in dealing with harassment because her attackers could simply create another account, some sending her messages such as *"new account up and running lol"* and *"It's great to be back after 30 seconds"*.[707]

Anonymous user profiles

4.65 Anonymity is an important and valued tool in protecting survivors of gendered violence, particularly those in marginalised communities, as well as for whistle-blowers and dissenting voices.[708] However, the evidence suggests that the ability to create anonymous user profiles also increases the risks of harm explored in this chapter.

4.66 Anonymity has been cited as one of the principal factors creating the 'disinhibition effect' when people do or say things online that they would not during physical interactions.[709] A 2017 study into the trolling of the McCann family reported that anonymous perpetrators,

---

[702] U.S Department for Justice (Morgan, R. and Truman, J.), 2022. Stalking Victimization, 2019. [accessed 28 September 2023].

[703] Yardley, E. 2021. Technology-facilitated domestic abuse in political economy: a new theoretical framework, *Violence against women,* 27 (10). [accessed 28 September 2023].

[704] Two qualitative studies from Australia found a high prevalence of harassment through creating multiple social media profiles in their samples of domestic abuse victims and survivors. The woman's ex-partner created a false profile, using pictures from when they were together, and sent messages to her friends to discredit her reputation. This left her fearing for her safety, leading her to shut herself off from her family and social media altogether. In another example, the woman would block her former partner on social media, but he would *"delete his whole account, not just deactivate, delete the whole account… so therefore that email address was no longer on the platform's system. Then he could go and make a new account and contact me again."* Source: eSafety Commission Australia, 2019. eSafety for Women from Culturally and Linguistically Diverse Backgrounds. [accessed 28 September 2023]; Dragiewicz, M., Harris, M., Woodlock, D., Salter, M., Easton, H., Lynch, A., Campbell, H., Leach, J. and Milne, L., 2019. Domestic Violence and communication technology: victim experiences of intrusion, surveillance and identity theft. [accessed 28 September 2023].

[705] Refuge, 2022. Marked as Unsafe. [accessed 28 September 2023].

[706] Kale, S, 2022. 11 years, 10 arrests, at least 62 women: how did Britain's worst cyberstalker evade justice for so long? *The Guardian,* 30 March. [accessed 28 September 2023].

[707] The writer and campaigner Caroline Criado Perez led a campaign to get more women on bank notes yet received intense backlash; at its peak receiving 100 to 200 tweets a minute, many of them abusive. Source: Criado Perez, C., New Statesmen, 2015. Caroline Criado-Perez's speech on cyber-harassment at the Women's Aid conference. The New Statesman, 27 September. [accessed 28 September 2023].

[708] McGlynn, C. and Woods, L., 2022. Violence against women and girls (VAWG) Code of Practice. [accessed 28 September 2023].

[709] Suler, J, 2004. The online disinhibition effect. [accessed 28 September 2023].

being unidentifiable as individuals, aligned themselves strongly with group identities and norms in ways that enabled trolling behaviour.[710]

4.67 Online abuse of public figures often comes from anonymous user profiles. A UNESCO study into the experiences of women journalists found that people identified by the women as 'unknown' or 'anonymous' constituted the highest-rated category of sources of online violence (57%).[711]

4.68 Evidence suggests that anonymity is related to the harassment of individuals, in particular, gender-based harassment. For example, an experiment set up in Israel in 2010 found that anonymous participants made more threats than identifiable participants.[712] And in a US study looking specifically at gendered harassment, it was found that perceptions of anonymity predicted intentions to engage in sexually harassing behaviours online.[713]

User networking

User connections

4.69 Functionalities that allow users to create online networks, such as user connections, can amplify abuse and harassment to a scale likely to cause significant fear and distress to victims and survivors.

4.70 A study involving victims and survivors of online harassment, alongside workers in trust and safety, explores how users connected to the perpetrator can amplify incidents of abuse and harassment. A moral accusation made to a network of users can trigger others within that network to individually send attacks, insults and in the worst cases, threats of death, rape and violence to the individually accused person. Users with many connections can be particularly effective at amplifying abuse and harassment. This process is presented as a model called 'morally motivated networked harassment'. [714]

4.71 A perpetrator can leverage their connections to get visibility of a target's user profile. A study into online forums discussing partner surveillance found evidence of perpetrators creating fake accounts to connect with second and third-degree connections to the victim or survivor in order access content that was otherwise unavailable to the public and the perpetrator.[715]

---

[710] Synnott, J., Coulias, A. and Loannou, M., 2017. Online trolling the case of Madeleine McCann. [accessed 28 September 2023].

[711] UNESCO (Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S.), 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].

[712] Lapidot-Lefler, N. and Barak, N., 2012., Effects of anonymity, invisibility, and lack of eye-contact in toxic online disinhibition. *Computers in human behaviour*, 28(2). [accessed 28 September 2023].

[713] Ritter, 2014. Deviant Behavior in Computer-Mediated Communication: Development and Validation of a Measure of Cybersexual Harassment, Journal of computer-mediated communication, 19 (2). [accessed 28 September 2023].

[714] An example includes a user falsely accused of being a Russian disinformation theorist by a highly followed conspiracy theorist on Twitter. He describes: "*when someone has 25K Twitter followers, they pile on really quickly, and it sort of becomes, especially in this case where it's very conspiratorially-minded thinking, that the accusations and the allegations sort of start to compound and build up on each other.*" He described being 'really upset' and 'very alone' as a result. This example shows how user networks can amplify harassment to a level that causes fear and distress to victims. Source: Marwick, A, 2021. Morally Motivated Networked Harassment as Normative Reinforcement. [accessed 28 September 2023].

[715] Some posts provide step-by-step instructions for creating a believable fake profile and befriending accounts connected to targeted individuals, allowing the perpetrator to access content visible to 'friends-of-friends'. Strategies can be highly targeted. As one poster describes: "*In my neck of the woods there are a lot of local bars that have 1000+ friends and guess*

User tagging

4.72 Perpetrators can tag victims and survivors in posts containing harassing language or violent threats. Amnesty International reports on women in public positions having their usernames tagged incessantly in abusive and threatening messages, causing significant distress to those targeted.[716]

User communication

Livestreaming

4.73 Several studies suggest that there is a link between gaming services, where harassment, abuse and threats are well known to occur, and livestreaming – gamers in a 2021 study believed that it leads to more bullying behaviour.[717] Gameplay on gaming services can often be livestreamed, which can encourage users to reveal identifying characteristics in a live, ephemeral context where moderation is more limited. The unbalanced anonymity while livestreaming – where the streamer is seen by all viewers, but viewing and interacting with the streamer through, for example, comment or chat functions, can be done anonymously – can create a risk.[718] And if streamers change their username and channel name to escape bullying, perpetrators are able to re-identify victims and survivors based on their video, image or sound, and continue the harassment.[719]

Direct messaging

4.74 Direct messaging enables perpetrators to harass, stalk and threaten individuals in a targeted manner. Evidence suggests that the harassment of public figures often occurs though direct messaging. A UNESCO study into the experiences of women journalists found that 48% had been harassed by unwanted messages.[720]

4.75 Messaging functionalities are central to the perpetration of online stalking. The Suzy Lamplugh Trust found that texts or direct messages were the most common digital stalking behaviour. Threats were also commonly made via online digital communication (likely and predominantly via direct messaging).[721]

Commenting on content

4.76 Some services allow users to reply to or comment on posted content. As noted in the Hate chapter, this functionality can be used to send hateful content to an individual; this content could also be abusive and may amount to harassment where a user sends multiple hateful

---

what? Every one of those 1000+ friends has now given access to those 1000+ people that allow friends-of-friends to see their info". *[Note: Facebook and Google part funded this research through gifts].* Source: Tseng, E., Bellini, R., McDonald, N., Danos, M., Greenstadt, R., McCoy, D., Dell, N and, Ristenpart, T., 2020. The Tools and Tactics Used in Intimate Partner Surveillance: An Analysis of Online Infidelity Forums. [accessed 28 September 2023].

[716] For example, a UK political comedian described a 'pile-on' of violence and abuse against her following a media appearance on a television debate programme. She described how: "*after the debate he continued to be rude about me on Twitter. That reached a whole new level. In the following 48 hours, I received 165 pages* of *Twitter abuse. Suddenly it went insane. In that, there were four or five death threats, rape threats, and things like that."* Source: Amnesty International, 2018. Online Violence against Women. [accessed 28 September 2023].

[717] See 'Risk factors: functionalities and recommender systems' section for more information. Source: McLean, L. and Griffiths, M. D., 2018. Female Gamers' Experience of Online Harassment and Social Support in Online Gaming: A Qualitative Study, *International Journal of Mental Health and Addiction*, 17(970-994). [accessed 28 September 2023].

[718] Zhou, Y. and Farzan, Y., 2021. Designing to stop live-streaming cyberbullying. [accessed 28 September 2023].

[719] Zhou, Y. and Farzan, Y., 2021.

[720] UNESCO (Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S.), 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].

[721] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].

or abusive comments to an individual. Harassment can also occur where a user sends multiple comments to the same individual which are not hateful or abusive but are calculated to produce alarm or distress and are oppressive and unreasonable.

4.77 Evidence suggests that abuse and harassment via comments, affects a significant number of users. The Oxford Internet Survey in 2019 reported that 27% of respondents had seen cruel or hateful comments or images posted online.[722] A 2014 study by Pew Research found that 22% of internet users had been a victim of online abuse or harassment in the comments section of their uploads.[723] Between January and March 2023, YouTube removed more than 853 million comments from videos for violating its Community Guidelines. Of these, more than 44 million were for harassment or bullying.[724]

Posting content

4.78 Harassment and stalking can take the form of public humiliation, where content about an individual is posted. This can particularly affect female politicians, who have been shown to receive a higher proportion of abusive posts and messages than male counterparts.[725] Social media posts have also been identified as the most common trigger for harassment (46.7%) in a study on academics' experiences of online harassment.[726]

4.79 A study into gang-affiliated individuals in Chicago analysed responses to gang-related content. Several posts were interpreted as direct threats to violence. This interpretation was contingent on hyper-local context (interpretation of language, familiarity with the events, institutions, and experiences noted in the text). The study concludes that the ability to post content on social media services allows individuals to broadcast threats to violence, not just with keywords but with mentions of offline events, people, local institutions and situations.[727]

Posting or sending location information

4.80 A 2021 UK study on social media users found that 9.5% of users surveyed reported "*tracking someone through GPS*".[728] This shows the pervasiveness of a behaviour that, in certain contexts or in conjunction with other behaviours, could amount to perpetration of cyberstalking. A study on discussion forums used by perpetrators suggests this location

[722] The Alan Turing Institute (Vidgen, B., Margetts, H. and Harris, A.), 2019. How much online abuse is there? [accessed 28 September 2023].
[723] This US-based study from Pew Research includes single occasions as measures of harassment, although in the UK the harassment offence is a course of conduct occurring on two or more occasions. Source: Pew Research, 2014. Online Harassment. [accessed 28 September 2023].
[724] YouTube, 2022. YouTube Community Guidelines enforcement – Google Transparency Report. [accessed 25 August 2023].
[725] A study on tweets sent directly to US candidates during the 2020 US election found that 15%-39% of all female candidates' tweets were abusive, compared to 5-10% percent of male candidates. Source: Institute for Strategic Direction (Guerin, C. and Maharasingam-Shah, E.), 2020. Public Figures, Public Rage: Candidate abuse on social media. [accessed 28 September 2023].
[726] Gosse, C., Veletsianos, G., Hodson, J., Houlden, S., Dousay, T.A. and Lowenthal, P.R., Hall, N., 2020. The hidden costs of connectivity, Learning, media and technology, 46(3). [accessed 28 September 2023].
[727] Patton, D. U., Pryooz, D., Decker, S., Frey, W. R. and Leonard, P., 2019. When Twitter Fingers Turn to Trigger Fingers: a Qualitative Study of Social Media-Related Gang Violence. [accessed 28 September 2023].
[728] Gunn, R., Tzani, C., Ioannou, M., Synnott, J. and Fumagalli, A., 2021. Cyberstalking among social media users: Perceptions, prevalence and characteristics. [accessed 28 September 2023].

tracking might be carried out through spyware devices or apps.[729] Some U2U services host geo-tagging functionalities[730] and this information can be communicated to users.

4.81 There is limited evidence of the nefarious use of location information, but it could potentially be used to commit or facilitate stalking offences, by enabling stalking activities to move offline. As noted in the Controlling or coercive behaviour chapter, the visibility of location information can also be used to track the location of a victim or survivor of previous abuse, including domestic abuse.[731] On geosocial online dating services users share identifiable images alongside location information with other users, creating the risk of harassment, stalking and other abuses if this information is recorded, retained, screenshotted or saved.[732]

Re-posting or forwarding content

4.82 User-generated content can be decontextualised and reposted, which results in 'context collapse'.[733] A study into academic professionals shows how cutting and re-posting content in unintended contexts can enable harassment. The study gives the example of a Princeton university professor, who gave a commencement speech in which she described the then US President as racist and sexist. The speech was recirculated on news sites and on social media services, and the professor received many threats to her work email – most of them racist and sexually violent. This harassment began when the video was shared with a different audience to the one it was originally shared with, just as posts that are re-posted are.[734]

Recommender systems

Content and network recommender systems

4.83 Recommender systems may amplify the risk of services being used to commit or facilitate harassment and threats. For instance, they can play a role in online 'pile-ons' in which large numbers of users 'attack' one other user, and which can often involve increasingly aggressive and abusive messages, and threats.[735] Recommender systems are likely to contribute to the virality of a pile-on – promoting the content that has attracted such rapid (negative) engagement to more users, and potentially specifically to those users more likely to engage with it and add to the abuse the victim receives. In this scenario, user

---

[729] *[Note: Facebook and Google part funded this research through gifts].* Source: Tseng, E., Bellini, R., McDonald, N., Danos, M., Greenstadt, R., McCoy, D., Dell, N., Ristenpart, T. 2020. The Tools and Tactics Used in Intimate Partner Surveillance: An Analysis of Online Infidelity Forums. [accessed 28 September 2023].

[730] Geotagging is the process of adding location data to media such as photos and videos, such as the coordinates of where a photograph or video has been taken. This occurs on most smartphones and tablets. Source: Paladin, n.d. Cyber and Digital Safety: are you a victim of cyberstalking? [accessed 28 September 2023].

[731] Woodlock, D, 2017. The Abuse of Technology in Domestic abuse and Stalking. [accessed 27 September 2023]; Sugiura, L., Blackbourn, D., Button, M., Hawkins, C., Tapley, J., Frederick, B., Nurse, J. R. C. and Rahime, B. S., 2021. Computer Misuse as a Facilitator of Domestic Abuse. p.34 [accessed 27 September 2023].

[732] Waldman, A, 2021. Navigating Privacy on Gay-Oriented Mobile Dating Applications. [accessed 28 September 2023].

[733] Context collapse in this example describes a piece of content being taken out of context to an audience it was not intended for. Source: Marwick, A., Boyd, D. 2010. I tweet honestly, I tweet passionately: Twitter users, context collapse and the imagined audience. [accessed 28 September 2023].

[734] Thrasher, S.W, 2017. Yes there is a free speech crisis. But it's victims are not white men. T*he Guardian*, 5 June. [accessed 28 September 2023]; Gosse, C., Veletsianos, G., Hodson, J., Houlden, S., Dousay, T.A., Lowenthal, P.R. and Hall, N., 2020. The hidden costs of connectivity: nature and effectis of scholars' online harassment, Learning, media and technology, 46(3). [accessed 28 September 2023].

[735] Thompson, J. D., Cover, R. 2021. Digital hostility, internet pile-ons and shaming: A case study. *Convergence: The International Journal of Research into New Media Technologies*, 28(6). [accessed 10 October 2024]

communication functionality and recommender systems work together to create this particular scenario.

4.84 The role of recommender systems in presenting users with preferred content and like-minded users also means they play a role in the spread of some types of content and groups of people who participate in harmful and illegal behaviour. One study into the 'Manosphere' – encompassing misogynistic movements such as the men's rights activists (MRAs), Men Going Their Own Way (MGTOW), and 'incels' among others – found that content recommender systems on social media redirected users who engaged with misogynistic videos to increasingly hateful, extreme, and inciting content.[736] The study noted that this content may radicalise users and connect them with Manosphere communities and ideologies that promote harmful and illegal behaviour towards women; in some cases, content associated with these communities encouraged and resulted in violent attacks.[737]

Risk factors: Business models and commercial profiles

4.85 There is limited evidence from investigative reporting that shows how online social media services may promote threatening and inciteful content by prioritising user engagement over user safety. Such content can have higher engagement rates, which services may be incentivised to promote if their business models seek to maximize user engagement at any cost.[738]

[736] Institute for Strategic Dialogue (Thomas, E. and Balint, K.), 2022. Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into the 'Manosphere.' [accessed 10 July 2024]
[737] Griffin, J., 2021. Incels: Inside a dark world of online hate, BBC News, 13 August. [accessed 11 July 2024]; Institute for Strategic Dialogue (Thomas, E. and Balint, K.), 2022. Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into the 'Manosphere.' [accessed 10 July 2024].
[738] Ellery B. and Mitib A., 2023. Social medic companies profit from misery spread by misogynistic influencers, The Times, 9 January. [accessed 29 July 2024].

5. Controlling or coercive behaviour (CCB)

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for controlling or coercive behaviour: how harm manifests online, and risk factors

Online services offer new ways for perpetrators to coerce and control partners, former partners, or their children. In England and Wales, an estimated 1.4 million women and 751,000 men aged 16 years and over experienced domestic abuse in the year ending March 2023, equivalent to approximately 5.7% of women and 3.2% of men.[739] The risks of harm from controlling or coercive Behaviour (CCB) are broad and can be life-threatening; they can affect a victim's or survivor's mental health, and affect their life in other ways, including their income. *Harm is often caused by the perpetrator's omnipresence in an individual's life.*

Common risk factors are present in the evidence; however, services should be aware that the offence is often perpetrated in complex and personal ways. CCB can also involve other offences, including threats and intimate image abuse (see the Harassment, stalking, threats and abuse chapter, and the Intimate image abuse chapter for further information).

Service type risk factors:

Social media services offer perpetrators multiple ways to monitor victims and survivors, and to pursue campaigns of targeted abuse. CCB often happens across *several social media services simultaneously.* **Messaging services** enable perpetrators to be a constant presence in the lives of victims and survivors.

Online dating services *and* ***online user-to-user pornography services*** *can also be used in cases of coercive control, particularly if the abuse involves the sharing of intimate images.*

User base risk factors:

Due to its significance in coercive control and other offences affecting women in particular, user base demographics are included as a general risk factor in the risk profiles. This is partly because CCB sits within a wider culture of gendered violence and misogyny. **Gender** is a risk factor, with women being more commonly and more severely affected. In England and Wales, in the year ending March 2023, 73.5% of domestic abuse-related crimes were female, and between 2020 and 2022,

[739] Office for National Statistics, 2023. Domestic abuse victim characteristics, England and Wales: year ending March 2023. [accessed 20 November 2024]. Controlling or coercive behaviour is outlined as abuse in the Domestic Abuse Act 2021.

67.3% of domestic homicide victims were female.[740] **Young women,** as well as women from **minority ethnic and racial backgrounds,** appear to be most at risk. Research indicates that **disability** and low **socio-economic status** can increase risk among women. While our evidence suggests that **LGBTQ+ communities** may be more at risk of CCB, more research is needed into possible 'hidden groups', which include male victims and survivors.

Functionalities and recommender systems risk factors:

Several functionalities enable monitoring practices. The most prominent are **fake user profiles**, which perpetrators can use to impersonate victims and survivors, as well as other individuals, to gain access to the target's account, to monitor and harass victims and survivors. Sending abusive or threatening **direct messages** – sometimes incessantly and across multiple services – can cause fear and distress to victims and survivors and make them feel that the perpetrators have a constant presence in their lives.

User connections allow users to build online networks, both around the perpetrators, and the victims and survivors. These networks can extend perpetrators' ability to coerce and control victims and survivors, for example by creating an environment for public humiliation, or getting contacts to join in with monitoring or harassment. Location tracking is also common in cases of CCB, so **posting or sharing location information** represents a risk factor for this offence.

Posting content gives perpetrators the ability to publicly post negative or personal information about victims and survivors, as well as to non-consensually share intimate images as part of campaigns of abuse.

Introduction

5.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the controlling or coercive behaviour (CCB) offence listed under the 'Relevant offences' header; and
- the use of these services for the commission and/or facilitation of this offence (collectively, the 'risks of harm').

5.2 We set out the characteristics of U2U services that we consider likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

[740] Office for National Statistics, 2023. Domestic abuse victim characteristics, England and Wales: year ending March 2023. [accessed 20 November 2024].

5.3 This chapter uses the term CCB primarily to reflect the offence listed in the Online Safety Act (the Act).[741] CCB specifically includes engendering psychological fear as a form of abuse.[742]

5.4 Sources referred to in this chapter may use other terms such as 'Tech abuse' (TA), 'Technology-facilitated abuse' (TFA), 'Technology-facilitated violence' (TFV) or 'Technology-facilitated Coercive Control' (TFCC or TCC), and we consider these to fall in scope of the legal definition of CCB. The language of coercive control is also used to explore how domestic abuse manifests online, based on the understanding that domestic violence is coercive, controlling, and profoundly contextualised in relationship dynamics, cultural norms, and structural inequality.[743] To stay aligned with the evidence, we use the terminology from the research when citing studies in this chapter.

5.5 This chapter will examine behaviour and content similar to that which is discussed in the chapters 'Harassment, stalking, threats and abuse', and 'Intimate image abuse', and a case of CCB is often made up of multiple offences. In this chapter, we have ensured that as much evidence as possible refers to these behaviours only when examined in a controlling or coercive context, to avoid duplication.

5.6 This chapter is focused on perpetration via U2U services. However, the literature on technology-facilitated domestic abuse and digital CCB often explores a wider range of technologies used to stalk, harass, threaten, and abuse partners, ex-partners or children. Indeed, most cases of perpetration of CCB on social media (94%) happen alongside other forms of technologically-enabled or offline domestic abuse, such as SMS messaging and spyware technology.[744]

5.7 Evidence relating specifically to U2U services will be used where available, although some of the evidence in this chapter may include the perpetration of CCB or similar behaviour through a wider range of technologies, partly because the evidence of harm directly tied to perpetration via U2U services is limited. Where such evidence has been included, it is to help services better understand CCB.

Relevant offences

5.8 The Act requires Ofcom to consider the risks of harm connected with specific offences. Regarding CCB, Ofcom is required to consider the risks of harm connected with the priority offence listed in Schedule 7 of the Act, being the offence of controlling or coercive behaviour in an intimate or family relationship.[745]

---

[741] An offence under section 76 of the Serious Crime Act 2015 (controlling or coercive behaviour in an intimate or family relationship).
[742] Stark, E., 2009. Coercive control: the entrapment of women in personal Life. [accessed 21 September 2023].
[743] For example, see: Dragiewicz, M., Burgess, J., Matamoros-Fernández, A., Salter, M., Suzor, N. P., Woodlock, D., Harris, B., 2018., Technology facilitated coercive control: domestic abuse and the competing roles of digital media platforms. [accessed 21 September 2023]; Harris, B. and Woodlock, D., 2022. Digital coercive control: Insights from two landmark domestic abuse studies. *The British Journal of Criminology*, 59(3). [accessed 21 September 2023].
[744] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].
[745] Section 76 of the Serious Crime Act 2015.

5.9 Controlling or coercive behaviour occurs where the victim-survivor[746] and the perpetrator are personally connected, the perpetrator repeatedly or continuously engages in behaviour that is controlling or coercive, and this behaviour has a serious effect on the victim-survivor, putting them in fear of violence or causing serious alarm or distress which has a substantial adverse effect on their usual day-to-day activities.[747]

5.10 Coercive behaviour can be an act (or a pattern of acts) of assault, threats, humiliation and intimidation or other abuse that is used to harm, punish, or frighten those in intimate or family contexts. Controlling behaviour includes a range of acts designed to make a person subordinate and/or dependent by isolating them from sources of support, exploiting their resources and capacities for personal gain, depriving them of the means needed for independence, resistance and escape, and regulating their everyday behaviour.[748]

5.11 Patterns of behaviour that amount to CCB can include harmful acts that often encompass other offences. A case of CCB might include cyberstalking, harassment and threats of violence, intimate image abuse, hatred towards minorities and other protected groups, and child sexual exploitation and abuse. This chapter links relevant evidence from across the Register with additional evidence to provide a comprehensive view of CCB and how it manifests online. Evidence relevant to other chapters will be cross-referenced.

5.12 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of the offence.

5.13 For more details on how services can assess whether content amounts to priority illegal content, and general illegal content, refer to the Illegal Content Judgements Guidance('ICJG').

How controlling or coercive behaviour manifests online

5.14 This section provides an overview on how CCB manifests online, and how individuals may be at risk of harm.

5.15 Manifestations of CCB can consist of many different types of content and activity happening at the same time or over a prolonged period. CCB can take place over many years in online and offline environments simultaneously. Although specific behaviours referred to in this chapter are also covered in other chapters, it is important to note that when occurring in a controlling and coercive context, alongside other harms, and often alongside offline abuse, the risk of harm for the victim-survivor is greater.

5.16 Controlling or coercive behaviour as covered in the Act includes engendering psychological fear as a form of abuse. Behaviours commonly covered include:

[746] The phrase "victims and survivors" is used as many survivors find the word "victim" disempowering. However, the phrase "victim-survivors" also encompasses those who did not survive CCB.
[747] It is important to note that the experience of partners is significantly more reported on than those of children in domestic abuse contexts. This is reflected in the evidence base available for this chapter. Research is limited on other groups such as children, male victims, minority ethnic groups, disabled people and LGBTQ+ people. This should be considered when reading this chapter.
[748] Crown Prosecution Service, 2023. Legal Guidance: Controlling or Coercive behaviour in an intimate or family relationship. [accessed 5 September 2023].

5.17 Monitoring, stalking or controlling behaviours, such as:

- **Device and app control:** includes a range of acts which are part of perpetrators' controlling behaviours of victims and survivors. This might also include 'unauthorised access' to a person's online accounts, by guessing passwords, or hacking.[749]
- **Public disclosure of private information and doxxing:** consists of the nonconsensual publication of private information online such as a person's address, phone number, driver's licence or other personal documents or personal information.
- **Impersonation, including catfishing:** is the use of digital technology to assume the identity of a person to access private information, exploit, embarrass, discredit or shame them, contact or mislead them, or create fraudulent documents. This can also include the online theft of documents, such as digital passport information and/or immigration documents.[750]
- **Psychological or emotional abuse or threats:** this comprises content and activity covered in the Harassment, stalking, threats and abuse chapter, including sexual and/or dating harassment behaviours.
- **Sexual abuse:** This includes behaviours covered in the Intimate image abuse chapter, Extreme pornography chapter, and Sexual exploitation of adults chapter.

5.18 This chapter will discuss how content and activity that amounts to a controlling or coercive context manifests online across these areas. To avoid repetition, we will not cite evidence from chapters previously mentioned, but services should consider that content and activity within those chapters may occur in a CCB context.

5.19 Controlling or coercive behaviour online often occurs in the context of an existing offline coercive and/or controlling relationship. A survey from Refuge found that 16% of women in the UK who had experienced abuse online said it was from a partner or former partner; this number rose to 22% for young women aged 18 to 34. One in five (21%) of those reporting tech abuse reported experiencing offline coercive control in addition to abuse online.[751] A similar survey in Australia found that one in four (25%) women aged 18+ said their most recent harmful experience online was facilitated by a current intimate partner; 16% said it was from a former partner, and 12% from a family member.[752] A 2013 survey by Women's Aid[753] of victims and survivors of domestic abuse found that 45% had experienced abuse online during their relationship.[754]

5.20 Controlling or coercive behaviour consists of many harm types, which often co-occur. In Refuge's 2022 survey, women who had experienced tech abuse from a current or former partner reported experiencing sexual harassment (29%); abusive or upsetting content being shared with them (25%); stalking and monitoring (23%); physical or sexual threats (22%);

---

[749] CIGI (Dunn, S., Vaillancourt, T. and Brittain, H.) 2023. Supporting Safer Digital Spaces. [accessed 18 November 2024].
[750] Henry, N., Vasil, S., Flynn, A., Kellard, K. & Mortreux, C. 2021. Technology-Facilitated Domestic Violence Against Immigrant and Refugee Women: A Qualitative Study. Journal of Interpersonal Violence. [accessed 05 September 2024]
[751] Refuge, 2021.
[752] Powell, A., Flynn, A., & Hindes, S. 2022. Technology-facilitated abuse: National survey of Australian adults' experiences. [accessed 26 November2024].
[753] Women's Aid is a charity working to end domestic abuse against women and children.
[754] Women's Aid (Laxton, C.), 2014. Virtual World, Real Fear: Women's Aid Report into Online Abuse, Harassment and Stalking. [accessed 23 September 2023].

non-consensual intimate image-sharing (17%) or threats to do so (14%); unauthorised access to their online accounts (21%); doxxing (11%) and deepfakes (4%).[755] [756]

5.21 Reflecting on their own experiences, victims and survivors report that motivations for coercive or controlling abusive behaviour towards them was 'a lot of the time' intended to intimidate the victim (93%), control the victim (93%), cause distress (87%), cause fear for safety (86%) and to isolate the victim or restrict their activities (82%).[757]

5.22 Controlling or coercive behaviour can take place over a long period, and severity can escalate over that time. 24% of women surveyed by Refuge said they experienced tech abuse on many occasions. On average, those who reported experiencing tech abuse said it had taken place for at least six months, and 48% said that the abuse they experienced on social media got worse over time.[758]

5.23 In addition, CCB spans on and offline spaces, with one in five victims supported by Refuge's Tech Abuse team reporting that their location had been compromised because of abuse taking place on online services.[759] Studies have found that tech-facilitated stalking and abuse are risk factors for domestic homicide for both women and their children.[760] Researchers at the University of Kent identified that 59% of domestic homicide review cases in a sample presented some evidence of digital activities or behaviours constituting cyberstalking; 17% of cases involved use of social media to surveil, send threats or intimidate victims; 12% of cases involved hacking or unauthorised access.[761]

5.24 Online, particularly user-to-user, services have shifted the landscape of domestic abuse by better enabling perpetrators to coerce and control victims and survivors at a distance.[762] Support services reflect that *"technology affords perpetrators a variety of ways to invade every aspect of women's lives at any time of day and night and from a distance"*.[763]

5.25 Over half (51%) of women who experienced abuse online from a current or former partner said that a third party was also involved in the abuse, such as the family or friends of the partner or ex-partner.[764] In addition, 50% of victims said their own family or friends had been targeted as part of the abuse, and 12% said their children had been targeted.[765]

---

[755] Refuge, 2021.
[756] Support services in Australia reported similar kinds of abuse experiences by victim-survivors they engage with, with 28% of women responding to the survey reporting experiencing multiple types of abuse from the same perpetrator: Powell, A., Flynn, A., & Hindes, S. 2022.
[757] Powell, A., Flynn, A., & Hindes, S. 2022.
[758] Powell, A., Flynn, A., & Hindes, S. 2022. Technology-facilitated abuse: National survey of Australian adults' experiences. [accessed 26 November2024].
[759] Refuge, 2021.
[760] Woodlock, D. et al, 2020.
[761] Todd, C., Bryce, J & Franqueira, V. N. L. 2020. Technology, Cyberstalking and Domestic Homicide: Informing Prevention and Response Strategies. *Policing and Society*, accessed [15.09.2024].
[762] Refuge, 2021.
[763] Refuge, 2021; Woodlock, D. et al, 2020.
[764] Refuge, 2021.
[765] Refuge, 2021.

5.26 As with offline coercive and controlling behaviour, and particularly domestic abuse cases, adult victims and survivors of online-facilitated CCB are overwhelmingly women, while perpetrators – whether current or ex-partners – are men.[766]

5.27 Certain forms of CCB are enabled by the perpetrator having a greater knowledge of, and ability to misuse, technology within current or former intimate and family relationships. Perpetrators might purchase, set up or maintain accounts on behalf of a partner or family member on U2U services (for example in online gaming services), and can take advantage of the increased access and knowledge imbalance that results from this. This can enable them to control, coerce, harass or abuse victims and survivors in a multitude of ways including unauthorised access, compromising account controls, and impersonation.[767]

5.28 Research has also found that in some online communities and forums, perpetrators of tech abuse share tactics and tools to monitor and abuse victim-survivors, including the creation of fake profiles to use second and third-degree connections to access a victim's profile.[768]

5.29 A survey of children and young people by charity Women's Aid found a link between potentially harmful social media content and unhealthy perceptions of, and attitudes towards, relationships.[769] When surveyed, children and young people who reported being exposed to misogynistic content[770] on social media were significantly more likely to agree with statements reflecting unhealthy perceptions of relationships compared to those who had not seen this kind of content. For example, those exposed to misogynistic content were more likely to agree that "hurting someone physically is okay if you say sorry after hurting them" (19%) compared to those not exposed (4%).[771]

5.30 Low incidences of reporting complicate the measurement of CCB. Like related offences, CCB is consistently under-reported.[772] Refuge found that half of victims and survivors (49%) said they told nobody about the abuse, with only 13% of women reporting the abuse to the social media platform where the abuse happened. Only one in ten victims and survivors (10%) reported it to the police. One in five victims and survivors (18%) did not tell anyone because they were not sure how to report the abuse.[773] Evidence also suggests that women

---

[766] In the Australian support service survey, respondents reported that perpetrators of tech abuse, in their experience, were most commonly men and boys, with a vast majority reporting that men (83%) and boys (35%) were using tech abuse to target victims 'a lot of the time', across both intimate partner, family member and/or acquaintance contexts: Flynn et al. (2021) Stakeholder survey.

[767] Freed, D. et al. (2018) ''A stalker's paradise': How Intimate Partner Abusers Exploit Technology', *ACM Conference on Human Factors in Computing Systems* (CHI 2018), accessed [05.09.2024].

[768] Tseng, E., Bellini, R., McDonald, N., Danos, M., Greenstadt, R., McCoy, D., Dell, N., Ristenpart, T., 2020. The Tools and Tactics Used in Intimate Partner Surveillance: An Analysis of Online Infidelity Forums. [accessed 21 September 2023].

[769] Women's Aid. 2023. Influencers and Attitudes. [accessed 4 September 2024].

[770] In the previously referenced Women's Aid survey (Women's Aid, 2023) the questions refer specifically to 'Andrew Tate content', due to the relevance of this influencer at the time of fieldwork, and recommend taking findings around this content to be representative of some types of existing online misogynistic content.

[771] Women's Aid, 2023.

[772] A lack of reporting can be compounded by insufficient data recording practices; the National Police Chiefs' Council (NPCC) highlights the challenges in assessing the scale of technology-enabled violence against women and girls (VAWG), due to data recording practices. It notes that while on average 10% of VAWG offences are recorded as occurring online, this is likely to be an underestimate, as most VAWG offences are likely to have a digital component. National Police Chiefs Council, 2023. Violence Against Women and Girls: Strategic Threat Risk Assessment 2023. [accessed 21 September 2023].

[773] Refuge, 2021.

may not identify as victims and survivors if asked directly; responses are higher if women are given specific examples of abuse to relate to.[774]

Risks of harm to individuals presented by controlling or coercive behaviour

5.31 User-to-user services provide unique methods for perpetrators to humiliate, manipulate or harass victims and survivors. Online CCB represents changing patterns and practices of domestic abuse, generating distinctive risks and outcomes.[775]

5.32 Themes emerging from the literature include the role that technology can play in enabling perpetrators to always be in contact with their victim, and the constant threat of harmful behaviours when facilitated remotely via technology —which causes hypervigilance in victims and survivors. Isolation from support systems, friends and family, and significant effects on health and wellbeing are commonly reported.[776] Nearly all victims and survivors of online CCB report experiencing harmful outcomes of varying forms and severity.

5.33 Some behaviours replicate or extend the dynamics of offline coercion; for example, sending abusive messages. Others present more online-specific methods of CCB; sharing or threatening to share intimate images online without consent, and hacking victims and survivors' accounts, enabling additional forms of coercion and control.[777] Account hacking, for example, can enable perpetrators to check victims' and survivors' correspondence and make sure they are at the location where they claim to be[778], as well as accessing and potentially sharing intimate images from that individual's account.[779]

5.34 The use of social media is now very common in cases of domestic abuse. All the most common forms of technology-facilitated domestic abuse identified by Refuge can be facilitated or committed by in-scope U2U services. These include online harassment and impersonation, threats of physical violence, and sharing (or threatening to share) intimate images or videos without consent.[780]

---

[774] Refuge reports (Refuge, 2021) that while more than one in three women (36%) reported experiencing forms of online abuse when asked about specific examples, only one in five (21%) self-identified as experiencing abuse on an online platform. Similarly, an Australian study found that many victims and survivors of intimate partner stalking do not identify stalking behaviour as such: Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

[775] Harris, B. and Woodlock, D., 2022.

[776] Woodlock, D. 2017. Fernet, M. Lapierre, A., Hébert, M. and Cousineau, M.-M., 2019. A systematic review of literature on cyber intimate partner victimisation in adolescent girls and women, *Computers in Human Behaviour* 100. [accessed 21 February 2024].

[777] Refuge found that 47% of victims and survivors reported that someone had access to their social media account against their wishes: Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[778] Refuge, 2022.

[779] An Australian study (Woodlock, D. 2017) reports a perpetrator hacking a victim's and survivor's account, before sharing intimate images with male contacts. Refuge reports (Refuge, 2021) a case in which 'Laurel's' partner accessed her social media accounts and impersonated her online, while intercepting and deleting messages to make her question her memory. Laurel was also physically abused by her partner, but spoke about the tech abuse and gaslighting as being the worst part of her experience.

[780] The most common forms of CCB online identified by Refuge are (Refuge, 2021): online harassment; stalking, monitoring and location-tracking; threats of physical and sexual violence; having accounts hacked or controlled; online impersonation; sharing of intimate images or videos without consent, or threats to share; having personal details shared online without consent, also known as 'doxxing'.

5.35 Women are disproportionately affected by CCB.[781] Offline domestic violence is primarily perpetrated by men with the aim of controlling or coercing their former or current partner.[782] A significantly greater proportion of women (than men) who had experienced tech abuse behaviours reported that it was perpetrated by a male partner or former partner.[783] Women were also significantly more likely than men to report having experienced multiple co-occurring harms, from the same perpetrator, at the time of their most recent experience of online abuse.[784]

5.36 LGBTQ+ populations[785] are also more likely, in relative terms, to have experienced partner and family abuse than their heterosexual equivalents. They are also more likely to suffer severe outcomes as a result. [786] [787] For example, a study in 2017 found that LGBT+ victims and survivors were almost twice as likely to have attempted suicide, or self-harmed, were more likely to have faced abuse from multiple perpetrators and were twice as likely to have experienced historic abuse from a family member. [788]

5.37 The risk of online CCB needs to be considered within the wider context of cases of coercive control. Online CCB can increase the risk of physical harm, as well as more directly causing psychological harm. Psychological harm is almost universal; the evidence consistently reveals emotional turmoil, life complications, and helplessness among victims and survivors.[789] A review of cyber intimate partner violence studies found a range of negative effects on victims: anxiety and depressive symptoms, psychological distress, isolation, social phobia, perception of loss of control over one's life, suicidal ideation and suicidal attempts.[790] Refuge found that 95% of women experiencing abuse on social media from a partner or former partner said the experience affected their mental health, or impacted them in other life-debilitating ways, such as by affecting their income.[791] More than one in three women felt anxious (37%) and stressed (36%), one in five felt ashamed (21%) and isolated (19%), while one in ten felt suicidal because of the abuse.[792] Researchers have also

[781] Office for National Statistics, 2018. Women most at risk of experiencing partner abuse in England and Wales: years ending March 2015 to 2017. [accessed 16 September 2024]
[782] In the year ending March 2023, the victim was female in 73.5% of domestic abuse-related crimes: Office for National Statistics, 2023. Domestic abuse victim characteristics, England and Wales: year ending March 2023. [accessed 20 November 2024].
[783] Powell, A., Flynn, A., & Hindes, S. 2022. Technology-facilitated abuse: National survey of Australian adults' experiences. [accessed 26 November2024].
[784] Powell, A., Flynn, A., & Hindes, S. 2022.
[785] As noted by SafeLives, a UK charity dedicated to ending domestic abuse, it is important to understand that different parts of this community can experience abuse in different ways: SafeLives, 2021. Transgender victims' and survivors' experiences of domestic abuse. [accessed 18 November 2024].
[786] Bisexual women were found to be 3 times more likely to experience abuse as heterosexual women: SafeLives, 2021. Bisexual victims and survivors' experiences. [accessed 16 September 2024].
[787] Office for National Statistics, 2018.
[788] Office for National Statistics, 2018.
[789] Brown, M. L., Reed, L. A., Messing, J. L., 2018. *Technology-Based Abuse: Intimate Partner Violence and the Use of Information Communication Technologies* in Ryan Vickery, J., Everbach, T. (eds). #NastyWomen: Reclaiming the Twitterverse from Misogyny. [accessed 21 September 2023].
[790] Fernet, M. Lapierre, A., Hébert, M. and Cousineau, M.-M., 2019.
[791] Refuge, 2021.
[792] Refuge, 2021.

found that individuals who have received offensive or threatening messages are more than five times more likely to have had suicidal thoughts.[793]

5.38 Online CCB may also indicate a risk of physical violence and loss of life.[794] Studies have found that tech-facilitated stalking and abuse are risk factors for domestic homicide for both women and their children.[795] The Chair of the Association of Police & Crime Commissioners, Vera Baird QC, has commented on a "*misconception about technology-facilitated abuse […] that online harassment is not real abuse – yet much of the abuse to which the victim is exposed is often tied to offline behaviours, including stalking and assault*".[796]

5.39 Research by Refuge also shows that women fear for their physical safety following online CCB. Almost one in five (17%) said they felt afraid of being attacked or subjected to physical violence following tech abuse.[797] 15% felt their physical safety was more at risk, 5% felt more at risk of 'honour'-based violence[798], and 12% felt afraid to leave the house because of online abuse.[799] Fear of physical harm is likely to contribute to the near-universal psychological harm. Victims and survivors have also reported having to relocate due to safety concerns for themselves and family members.[800]

5.40 Some victims and survivors are left feeling unsafe online.[801] If victims and survivors disengage from online services because of CCB, this can have significant adverse effects including isolating them from family, friends, professional and social networks (thereby reducing their ability to access support). Being inaccessible online can in fact heighten abuse or amplify the risk of physical contact to enable the perpetration of abuse.[802]

5.41 All gendered forms of online abuse, including coercive and controlling behaviour, also have a 'chilling' effect on women and girls' freedom of speech. Refuge found that 38% (nearly 2 in 5) of women who experienced abuse on social media from a partner or former partner said they felt unsafe or less confident online as a result.[803] At a societal level this results in the silencing and exclusion of women and girls from online spaces.

5.42 Research found that risk of online CCB appear to increase once victims and survivors have separated from their partners, known as 'post-separation'. Relationship breakdown can be

---

[793] McManus, S., Bebbington, P.E., Tanczer, L., Scott, S. and Howard, L.M. 2021. Receiving threatening or obscene messages from a partner and mental health, self-harm and suicidality: results from the Adult Psychiatry Morbidity Survey. [accessed 18 November 2024].

[794]Woodlock, D., McKenzie, M., Western, D. and Harris, B., 2020. Technology as a Weapon in Domestic abuse: Responding to Digital Coercive Control. *Australian Social Work,* 73 (3). [accessed 21 September 2023].

[795] Woodlock, D. et al., 2020.

[796] All-Party Parliamentary Group on Domestic Violence: as referenced in Women's Aid, 2017. Tacking domestic abuse in the digital age. [accessed 21 September 2023].

[797] Refuge, 2021.

[798] Honour-based abuse is a crime or incident committed to protect or defend the 'honour' of a family or community: Metropolitan Police, n.d. What is honour-based abuse?. [accessed 5 September 2023].

[799] Refuge, 2021.

[800] Henry, N., Vasil, S., Flynn, A., Kellard, K. & Mortreux, C. 2021.

[801] Refuge found (Refuge, 2021) that 38% of women who experienced abuse on social media from a partner or former partner said they felt unsafe or less confident online as a result.

[802] A domestic abuser service provider describes the risks associated with blocking perpetrators. *"You have to be really careful. You can't even really tell anyone to block anyone 'cause that could escalate things as well. It is literally a case-by-case basis. Some victims know. I'm keeping him sort of subdued by just taking his behaviours, but if I react then maybe he'll react."*: Sugiura, L., Blackbourn, D., Button, M., Hawkins, C., Tapley, J., Frederick, B., Nurse, J. R. C., Rahime, B. S., 2021. Computer Misuse as a Facilitator of Domestic Abuse. [accessed 21 September 2023].

[803] Refuge, 2021.

a common trigger[804]; 66% of women who experienced tech abuse from an intimate partner said they were an ex-partner at the time of the abuse, compared to 18% who said the perpetrator was a partner at the time of the abuse.[805]

5.43 Individuals can be coerced into exploitative situation by gangs and criminal organisations. While CCB is used to coerce and exploit individuals, the offence only covers coercion and control where the victim and perpetrator are personally connected; for example, if they are or have been in an intimate personal relationship with each other. As a result, organised criminal exploitation is not in scope of CCB offences. Some of the risk factors identified in this chapter are likely to be applicable. More information on exploitation can be found in the Sexual exploitation of adults chapter.

Evidence of risk factors on user-to-user services

5.44 We consider that the risk factors below are likely to increase the risks of harm relating to CCB. This is also summarised at the start of the chapter.

Risk factor: Service types

5.45 Research indicates that social media services can be used to commit or facilitate CCB. Messaging services, user-to-user pornography services, and dating services may also be risk factors.

Social media services

5.46 There is strong evidence that a significant proportion of online CCB takes place on social media services. Research by Refuge shows that 36% (over one-third) of UK women have experienced abuse on social media or other online services.[806] It also shows that 60% of women who reported experiencing abuse on social media services also reported emotional abuse, and 21% (more than 1 in 5) experienced coercive control.[807]

5.47 Social media services can be used in 'proxy stalking', a mechanism that allows perpetrators to monitor and contact victims through other people.[808] Young people may be more likely to use social media for coercive control, with some research showing that under-30s are more drawn to using services such as social media.[809]

Messaging services and dating services

5.48 The risk posed by messaging services is further supported by the potential use of direct messaging for controlling and coercive behaviour. The ability to send direct messages across multiple devices and services allows perpetrators to maintain a presence in the lives

[804] Woodlock, D. 2017.
[805] Refuge, 2021.
[806] Refuge, 2021.
[807] Refuge, 2021.
[808] An Australian study found multiple cases of perpetrators using the social media pages of shared friends, family or even their children for monitoring purposes: Woodlock, D. 2017.
[809] Younger persons are more likely to use functionalities associated with social media services (unauthorised access to accounts, creation of fake profiles), whilst older people are more likely to use physical covert devices: Sugiura, L. et al., 2021.

of victims. Direct messaging is central to messaging services, as well as many social media services and dating services.

5.49 Research shows that dating services may be used by offenders to manipulate and impersonate victims or share intimate content. A common example of perpetration on dating sites is perpetrators setting up fake profiles for their partners to divulge victim and survivors' personal information, share intimate images or engage in sexual conversations with other users.[810]

User-to-user pornography services

5.50 Research shows that perpetrators of domestic abuse share intimate images (reported by 17% of victims and survivors) or threaten to share intimate images (reported by 14% of victims and survivors) in order to retain control over their victim.[811] User-to-user pornography services host significant amounts of intimate image abuse; this and other services which are known to facilitate intimate image abuse are discussed in the Intimate Image Abuse chapter.

Risk factors: User base

User base size

5.51 As discussed in the user networking section, perpetrators use networks of individuals to monitor victims and survivors from afar, as well as to incite harassment through others. This suggests that services with a large user base present a higher risk, given that they are more likely to host more users in both a victim and survivor's and/or perpetrator's network.

5.52 WhatsApp (web and app) and the Facebook Messenger app are the most popular messaging services, reaching 89% and 59% of UK online adults in September 2024 respectively.[812] These services are the most-used in cases of technology-facilitated domestic abuse. Refuge found that 71% of victims and survivors had experienced abuse on Facebook and 53% on WhatsApp.[813] TikTok and Snapchat reached 53% and 22% of UK online adults respectively in September 2024.[814] These services are less commonly used in abuse cases, with 20% of victims and survivors reporting abuse via Snapchat, and 13% on TikTok.[815] On the basis of this evidence, we therefore consider that it is likely that incidences of CCB would be higher on services with larger user bases.

User base demographics

5.53 The following section outlines essential evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

---

[810] Sugiura, L. et al., 2021.
[811] Refuge, 2021.
[812] Ipsos, Ipsos iris online audience measurement service, (BG) WhatsApp (web and app) and (APP) Facebook Messenger (app only), September 2024, UK online adults aged 18+. Note that Facebook's messenger service can be accessed through the main Facebook site and app which is not included in this data.
[813] Refuge, 2022.
[814] Ipsos, Ipsos iris online audience measurement service, brand group (BG) TikTok, and (BG) Snapchat, September 2024, UK online adults aged 18+.
[815] Refuge, 2022.

5.54 While demographic information can be limited due to the sensitivity and importance of anonymity in CCB cases[816], our evidence suggests that users' **age, gender, race and ethnicity, sexual orientation, gender identity,** as well as **socio-economic status** and **disability** could lead to an increased risk of harm to individuals.

5.55 Age can be a relevant risk factor in cases of domestic abuse. Data from the Office for National Statistics identifies women aged 16 to 24 years old as most at risk of partner abuse, with risk decreasing for older cohorts.[817] Evidence suggests that abuse perpetrated on social media is particularly common among young women. Refuge reports a third (30%) of women aged 16 to 19 having experienced CCB in a relationship, rising to over half (51%) when presented with a list of potentially controlling or coercive behaviours.[818] One in four (26%) young women report having their social media accounts monitored by a partner or former partner, making it one of the most experienced forms of coercive control among young women.[819] Support service workers were most likely to report those aged 18 to 24 years and 25 to 34 years as victims of tech abuse 'a lot of the time' (73% and 71%), followed by those aged 17 or under and 35 to 44 years old (55% and 50%).[820]

5.56 Although evidence into CCB online specifically is limited, there is a wealth of evidence indicating that domestic abuse affects more women than men. Eighty-three percent of high-frequency victims and survivors (having experienced more than ten crimes) are women.[821]

5.57 Women were almost three times more likely than men to experience sexual abuse from the person who abused them online.[822] This suggests that women are more likely to experience online CCB as part of a continued pattern of abuse. Women were more likely to experience different types of co-occurring abuse from the same perpetrator of their most recent experience of tech-facilitated abuse, with 27.8% (more than 1 in 4) of all women and 18.6% (nearly 1 in 5) of all men surveyed reporting this specific dynamic.[823] Refuge reports that 21% (more than 1 in 5) of those who experienced online abuse on social media reported experiencing coercive control in addition to this.[824]

5.58 There is also evidence to suggest that CCB is disproportionately perpetrated by men. For those self-reporting tech abuse victimisation in the Australian national survey, a majority of victims reported that in their most recent victimisation experience the perpetrator was a

---

[816] A literature review on intimate partner violence found that the majority of studies lacked adequate information about demographic characteristics such as age and geographical region: Grimani, A., Gavine, A., and Moncur, W. 2022. An Evidence Synthesis of Covert Online Strategies Regarding Intimate Partner Violence, *Trauma, Violence and Abuse*, 23(2). [accessed 21 September 2023].

[817] Young women aged between 16 and 19 (7.6%) and 20 and 24 (7.4%) were significantly more likely to have experienced partner abuse in the 12 months before interview than women aged between 45 and 54 (5.6%) or between 55 and 59 (4.4%): Office for National Statistics, 2018.https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/womenmostatriskofexperiencingpartnerabuseinenglandandwales/yearsendingmarch2015to2017

[818] Refuge, 2022.

[819] Refuge, 2022.

[820] Powell, A., Flynn, A., & Hindes, S. 2021. Stakeholder survey. Technology-facilitated abuse: A survey of support services stakeholders. [accessed 26 November 2024].

[821] Walby, S. and Towers, J., 2018. Untangling the concept of coercive control: Theorizing domestic violent crime. Criminology and Criminal Justice, 18(1). [accessed 21 September 2023].

[822] Refuge, 2021.

[823] Stakeholder survey (2022).

[824] Refuge, 2021.

man (62.1%) compared to 31.1% saying the perpetrator was a woman.[825] For example, one study of 96 cases of domestic abuse recorded by the police found that men are more likely to be repeat perpetrators, and more likely than women to use physical violence, threats and harassment.[826] Data from Galop shows that male perpetration is also higher among LGBTQ+ communities. A survey of LGBTQ+ victims and survivors found that 71% of individual perpetrators identified as male and 29% as female.[827]

5.59 While women are at significantly higher risk of CCB, it is argued by some that acknowledgment of this should not obscure the experiences of men or those with other gender identities.[828] [829] Although the proportion of male compared to female victims and survivors of domestic violent crimes is relatively low, the number of individuals this represents is still substantial.[830]

5.60 Looking at the prevalence of domestic abuse more broadly among minority ethnic groups shows that some ethnic identities are more at risk than others. Certain minority ethnic groups were found to be less at risk than white women. ONS data found that women who identified with the mixed/multiple ethnic group (10.1%) were more likely to have experienced partner abuse in the past 12 months than any other ethnic group. Asian and Asian British women were the least likely to have been victims of partner abuse (2.8%). White women (6.5%) were twice as likely to have experienced partner abuse as Asian and Asian-British women.[831] Women from communities with certain conservative or religious norms can be particularly vulnerable, as technology might be used to threaten or shame a partner wishing to leave a relationship, for example through Intimate Image Abuse, meaning they might be at increased risk of 'honour violence'.[832]

5.61 Data from the Office for National Statistics suggests that lower socio-economic status can increase the risk of partner abuse among women. Women who live in households earning less than £10,000 a year were more than four times as likely (14.3%) to have experienced partner abuse in the past 12 months than women living in households with an income of £50,000 or more (3.3%), while women living in social housing were more likely to have experienced partner abuse in the past 12 months (11.1%) than private renters (7.8%) or owner-occupiers (4.1%).[833] Women in rural and remote areas can face additional factors including increased dependence on communication technologies for safety and connection,

---

[825] Powell, A., Flynn, A., & Hindes, S. 2022. Technology-facilitated abuse: National survey of Australian adults' experiences. [accessed 26 November2024].

[826] Hester, M., 2013. Who Does What to Whom? Gender and Domestic abuse Perpetrators in English Police Records. *European Journal of Criminology*, 10(5). [accessed 21 September 2023].

[827] From a sample of 626 LGBTQ+ victims and survivors based in Greater London. Galop, (Magić, J., Kelley, P.), 2018. LGBT+ People's Experiences of Domestic Abuse: a report of Galop's domestic abuse advisory service. [accessed 21 September 2023].

[828] Sugiura, L. et al., 2021.

[829] Furthermore, it is recognised that male survivors of domestic violence are often not able to access support or are not taken seriously due to a lack of recognition of the problem.

[830] Almost three-quarters (74%) of domestic violent crime victims were female and 82% of domestic violent crimes were against women in Crime Survey for England and Wales data reviewed from 2008 to 2013, but there were still 79,473 men experiencing 219,118 cases of domestic abuse in this sample: Donovan, C., and Barnes, R., 2021. Re-tangling the concept of coercive control: A view from the margins and a response to Walby and Towers (2018). *Criminology and Criminal Justice*, 21(2). [accessed 21 September 2023].

[831] Office for National Statistics, 2018.

[832] Brookfield, K. et al. (2024), Technology-Facilitated Domestic Abuse: An under-recognised safeguarding issue? *The British Journal of Social Work* 54.1, accessed [05 September 2024].

[833] Office for National Statistics, 2018.

and being part of smaller communities where reputational damage can have a greater effect than in other communities.[834]

5.62 Minority sexual identities are likely to be a risk factor for online CCB, based on the existing evidence. These populations can face specific forms of partner abuse.[835] Although online CCB is not measured specifically, both online abuse and intimate partner abuse are higher among LGBTQ+ populations. LGBTQ+ women are much more likely to have experienced online abuse than women who do not identify as LGBTQ+. Three in four (75%) LGBTQ+ female survey respondents said they had experienced online abuse, compared to 33% of non-LGBTQ+ women.[836] The ONS found that bisexual women were nearly twice as likely to have experienced partner abuse in the past twelve months as heterosexual women (10.9% versus 6.0%).[837]

5.63 Disability has been identified as a potential risk factor for online CCB. In a 2021 Australian study engaging with women with limiting and effecting conditions and with relevant support services, researchers found that women with disabilities can face greater risks and effects from technology-facilitated coercive control, as they may have less control over their online accounts and safety settings. Abuse can come from a wider range of perpetrators, including parents, carers and children.[838] Additionally, negative impacts can be more severe for those with disabilities who rely on technology to stay in touch with family, friends, and to access support services.[839]

5.64 Evidence linking online CCB and gender identity is also lacking. However, evidence of intimate partner violence more broadly indicates that being transgender is likely to be a risk factor. In a 2015 US-based survey, more than half of trans respondents (54%) had experienced some form of intimate partner violence, including acts of coercive control and physical harm.[840]

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

5.65 Perpetrators can gain access to victims and survivors accounts, and then impersonate them through the user profiles associated with those accounts. This results in fake user profiles, which perpetrators and their networks can also create to publicly humiliate their victims.

5.66 Analysis of media reports found that 'fake' user profiles are a common CCB tactic.[841] Perpetrators can impersonate victims and survivors through these profiles, as well as use them to monitor, harass or humiliate their target. They can also create fake user profiles

---

[834] Woodlock, D. et al, 2020.
[835] A national UK LGBTQI+ anti-violence charity Galop highlights the specific issues of partner abuse unique to the experiences of LGBTQI+ people, such as the threat of disclosure of sexual orientation and gender identity to family, friends, or work colleagues: Galop, n.d. Domestic Abuse. [accessed 5 September 2023].
[836] Refuge, 2021.
[837] Office for National Statistics, 2018.
[838] eSafety Commissioner (Harris, B., and Woodlock, D.), 2021. For my safety: experiences of technology-facilitated abuse among women with intellectual disability to cognitive disability. [accessed 4 September 2024].
[839] WESNET, 2022. How tech abuse affects women with disabilities. [accessed 4 September 2024].
[840] National Center for Transgender Equality, 2015. US Transgender Survey. [accessed 8 September 2023].
[841] Sugiura, L. et al., 2021.

that represent fictitious people or real people known to victims and survivors.[842] Refuge reports that 29% of victims and survivors have been impersonated.[843]

5.67 Stories from victims and survivors reported by Refuge demonstrate the potential scale of abuse through fake accounts and their associated fake user profiles. One individual reported being threatened by a former partner from fake accounts across services, with 40 accounts reportedly created.[844] Another reported blocking her former partner, only to find over 120 fake accounts created by him in the space of a few weeks to continue his harassment of her.[845] More detail on harassment through fake user profiles can be found in the Harassment, stalking, threats and abuse chapter.

5.68 Fake user profiles can be particularly difficult for victims and survivors to cope with because it can be unclear whether they are fake or represent the identities of real users. In a Refuge study, 26% (more than 1 in 4) of victims and survivors reported being contacted repeatedly by an account that they suspected to be fake.[846] Limited recourse exacerbates this issue, with Refuge reporting that services almost never take down suspected fake accounts and their associated fake user profiles.

5.69 As described in the Harassment, stalking, threats and abuse chapter, user connections and networks can be leveraged by perpetrators to create fake user profiles that facilitate CCB. A study into online forums discussing partner surveillance found evidence of perpetrators using second and third-degree connections to gain visibility of a target's user profile without connecting with them directly. For example, one post provides a step-by-step account of how to create a believable fake user profile and befriend users who are friends of the victim and survivor.[847]

User networking

User connections

5.70 Functionalities that allow users to build online networks such as user connections are a risk factor. The involvement of networks is common in cases of online CCB, as it allows perpetrators to use other people to monitor or contact the victims and survivors in a practice called 'proxy stalking'. This can create the impression that the perpetrator knows and has seen everything, which is a primary mechanism for coercion and control. For more than half of women (52.4%) who were abused on social media services by current or former partners, a third-party connection was involved in the perpetration of the abuse.[848]

---

[842] The study lists examples of this, such as a man who set up accounts on swingers' and dating accounts in a woman's name with her workplace listed to discredit her, or a woman setting up a fake account under her ex-partners name and sending abusive messages to herself, before reporting this to the police: Sugiura, L. et al., 2021.

[843] Refuge, 2022.

[844] Refuge response to Ofcom 2022 Call for Evidence: First phase of online safety regulation. This included the quotation from a target of CCB: "*When I was pregnant I was getting threats about my child. A lot of (the messages) were fake accounts – so it was over 40 accounts […] I reported three times. […] He'd send me voicemails - you can do that on [social media platforms]. He made other accounts where he threatened to kill me and then he messaged my family on [social media platforms]".*

[845] Refuge, 2022.

[846] Refuge, 2021.

[847] Tseng, E., Bellini, R., McDonald, N., Danos, M., Greenstadt, R., McCoy, D., Dell, N., Ristenpart, T., 2020.

[848] Melton, H, 2007. Stalking in the context of intimate partner abuse: In the victims' words. *Feminist Crimonolgy* 2(4). [accessed 21 September 2023].

5.71 In addition to using their own pre-existing networks comprising friends and family,[849] studies report that perpetrators are also using victims' and survivors' networks to intimidate, harass and humiliate women, or challenge women's accounts of abuse.[850] Victims' and survivors' networks are often not aware of how they are being used.

5.72 Access to user connections can also be used to target the network of the victim and survivor as part of the abuse. Refuge found that 19% indicated that abuse got worse over time because the perpetrator started targeting their family or friends.[851] While this statistic does not specify whether family and friends were targeted via social media, this is likely to have been the case, given the usefulness of social media services in compiling networks.[852]

User communications

Direct messaging

5.73 The ability to send direct messages is a risk factor for CCB. Perpetrators are often able to send direct messages across multiple devices and services, allowing them to have a constant presence in the lives of their targets. This is an important tactic in controlling or coercive contexts.

5.74 Direct messages, texts and phone calls often rapidly cycle between verbal abuse, threats of violence and self-harm, and threats of punishment for not responding. Many abusers deliberately use veiled references and avoid explicit threats, which makes it difficult for women to provide clear evidence of the abuse they are experiencing to police, courts, and telecommunications companies.[853] [854] In a UK-based Refuge study, an individual reported how a perpetrator contacted her "*professional and personal accounts with messages, hundreds of messages. If (my employer) posts anything on social media, he will comment on there*".[855]

5.75 An Australian study found that perpetrators could use their friends' and family members' devices and accounts to contact victims and survivors via direct messages. This means that while women sometimes blocked their abusers' number or account, or had orders prohibiting communication, it was not possible for them to block all possible sources of contact in their abuser's network.[856]

Posting content (text, images)

5.76 Although posting content is common across U2U services, our evidence points to this functionality on a service as being a specific risk factor in CCB.

---

[849] 19% of victims and survivors said that the family of their partner or ex-partner was involved in the abuse, and 8% said their partner's friends were involved: Refuge, 2021.
[850] Dragiewicz, M., Harris, M., Woodlock, D., Salter, M., Easton, H., Lynch, A., Campbell, H., Leach, J., Milne, L., 2019. Domestic Violence and communication technology: victim experiences of intrusion, surveillance and identity theft. [accessed 21 September 2023]; Grimani, A., Gavine, A. and Moncur, W. 2022.
[851] Refuge, 2021.
[852] Woodlock, D. 2017.
[853] Dragiewicz, M. et al., 2019.
[854] Problematic use of this functionality requires an understanding of context for it to be identified as CCB. Repeated direct messages, like frequently messaging one's partner to check their location, can be harmless or abusive, depending on the overall context of the relationship. Dragiewicz, M. et al., 2019.
[855] Refuge, 2021.
[856] Dragiewicz, M. et al., 2019.

5.77 There are a number of different kinds of content that can cause relevant harm if posted, especially in a public context. These include identifying information ('doxxing'), negative information, intimate images, and threatening words or images.

5.78 The ability to post content, combined with user networks, facilitates 'doxxing.' This describes sharing identifying information about a particular individual online with intent to cause harm or distress. Doxxing often causes harm by encouraging other users in their network to join in with the harassment of victims and survivors. Refuge report that 18% of victims and survivors had experienced doxxing.[857]

5.79 Posting content allows perpetrators to share negative information on open channels of communication about victims and survivors in cases of CCB. These services afford the perpetrator an audience where a victim or survivor can be tormented in view of their community and personal connections, like friends and family.[858] An Australian study reported that 33% of victims and survivors have had negative information about them posted on social media services.[859]

5.80 The ability to post content such as images also enables intimate image abuse (IIA). The evidence indicates a significant overlap between the offence of intimate image abuse (explored in full in the Intimate Image Abuse chapter) and CCB. The non-consensual sharing of intimate images to other users of a platform affects a significant proportion of victims and survivors of CCB. Refuge reports that 29% of victims and survivors experience intimate image abuse.[860] Domestic abuse is also one of the three types of intimate image abuse identified by the Revenge Porn Helpline, indicating significant overlap between CCB and intimate image abuse.[861] Of the 376 prosecutions for intimate image abuse offences recorded in the year ending March 2019, 83% (more than 4 in 5) were flagged as being domestic abuse-related.[862] Other studies provide case studies of intimate image abuse in the context of CCB.[863]

Posting or sending location information, user groups, user events, user tagging

5.81 The ability to post or send location information is an important risk factor for CCB. The sharing of location information – sometimes inadvertently – can facilitate offline stalking. In a 2021 UK study on cyberstalking (not specifically in the context of CCB), 9.5% (nearly 1 in

[857] The Refuge study (Refuge, 2022) also provides qualitative examples of doxxing in the context of domestic abuse. For example, 'Paula', whose former partner waged a campaign of harassment, publicly accused her of lying about the domestic abuse that she faced and encouraging others to abuse her. Direct threats of harm were made, and her name and address were publicly shared from the abuser's account.
[858] Woodlock, D. 2017.
[859] An individual described how her former partner publicly claimed she had given him a sexually transmitted infection – this information was read by her teenage son's friends, among other people. The same study found perpetrators publicly shaming victims and survivors as 'punishment' for transgressions. Practitioners report behaviours such as a 'status update' where the perpetrator blames his problems on the victims and survivors, calls them names and accuses them of shameful behaviour. This can result in 'comments' of support to him from family and friends, leaving victims and survivors feeling isolated and 'ganged up on' by an entire community. Woodlock, D. 2017.
[860] Refuge, 2022.
[861] Sharratt, E., 2019. Intimate image abuse in adults and under 18s. [accessed 21 September 2023].
[862] Office for National Statistics, 2019. Domestic abuse and the criminal justice system, England and Wales: November 2019. [accessed 8 September 2023].
[863] For example, in one case provided by the Law Commission, an ex-partner set up a fake Facebook account in their ex-partner's name and uploaded intimate images of her, which were then viewed and copied to pornography sites, where on one website the picture was viewed over 48,000 times: Sharratt, E, 2021.

10) of perpetrators reported 'tracking someone through GPS'.[864] This section draws on evidence from the Harassment, stalking, threats and abuse chapter. More detail on cyberstalking can be found in this chapter.

5.82 A recent survey by Refuge found that 41% (more than 2 in 5) of victims and survivors had experienced location tracking.[865] The use of online services such as social media services can generate a variety of information points regarding location. Functionalities that allow for the tagging of other users and locations can make the activities and whereabouts of the victims and survivors visible to potential perpetrators.[866] Perpetrators can use geolocation tracking (for example, attached to status updates) to see where their partners and former partners are.[867]

5.83 In addition, functionalities such as user groups and events can also be used to track victims' and survivors' locations. A study in Canada with students found that 11% of participants had experienced a former intimate partner turning up at an event they intended to go to, as posted on their Facebook account.[868]

Content editing

Editing visual media

5.84 The editing of visual media such as images and videos to create deepfakes,[869] which can then be shared on U2U services, is likely to feature in some CCB cases. Refuge found that 4% of victims and survivors had experienced deepfakes.[870] While no specific evidence exists for this, threatening to create deepfakes may also be present in some cases of CCB.

Recommender systems

Content recommender systems

5.85 Content recommender systems are commonly designed to personalise content, and users who positively engage (for example, liking, sharing, and commenting) with certain categories of content will be served more of that content. From this, it is understood that users who are inclined to engage with CCB and adjacent content are likely to see more of that content in their feed. Some research argues that recommender systems may suggest content that contains instructions and methods of controlling or coercing partners, such as how to hack accounts or use tracking devices. Providing this information to potential perpetrators can amplify the risk of CCB behaviour. A UK paper exploring how simple web searches facilitate domestic abuse concludes that *"algorithms need to be adapted […] to avoid directing perpetrators to guidance informing them as to how to hack into their*

---

[864] Gunn, R., Tzani, C., Ioannou, M., Synnott, J., Fumagalli, A., 2021. Cyberstalking among social media users: Perceptions, prevalence and characteristics. [accessed 21 September 2023].
[865] Refuge, 2022.
[866] Woodlock, D. 2017.
[867] Sugiura, L. et al., 2021.
[868] Chaulk, K., Jones, T. 2011., Online Obsessive Relational Intrusion: Further Concerns About Facebook. *Journal of Family Violence,* 26. [accessed 21 September 2023].
[869] Deepfakes are a specific type of media that involves the use of AI algorithms, particularly generative AI models, to modify videos, images or audio to create realistic synthetic content. This is often done by superimposing the face of a person onto the body of another person in a video or image as well as voice manipulation with lip syncing. Deepfakes are commonly shared as user generated content on user-to-user services but could also potentially be created using functionalities present on user-to-user services. Deepfake technology is currently used to create content that can be harmful; however, we acknowledge that it may also have positive use cases.
[870] Refuge, 2021.

partner's accounts or stalk partners".[871] Recommender systems may increase the risk of perpetrators coming across content that can be used for abusive purposes (such as spyware or information related to their partners or former partners) if the algorithm recommends content based on a perpetrator's previous search history or interaction with content. This, in turn, increases the risk of perpetrators finding content that enables them to commit abusive or controlling behaviour.

Risk factors: Business models and commercial profiles

5.86 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[871] Sugiura, L., Blackbourn, D., Button, M., Hawkins, C., Tapley, J., Frederick, B., Nurse, J. R. C., Rahime, B. S., 2021. Computer Misuse as a Facilitator of Domestic Abuse. [accessed 21 September 2023].

6. Intimate image abuse

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for intimate image abuse: how harms manifest online, and risk factors

This chapter looks at offences relating to non-consensually sharing or threatening to share intimate images. Between 2019 and 2022, 24 police forces recorded a total of 13,860 intimate image offences, with more offences recorded in the first six months of 2022 than in all of 2020.[872]

Such acts can have serious negative impact on individuals, causing mental health issues, with considerable distress and anxiety experienced by victims and survivors. This can include shame, helplessness, self-blame, isolation and humiliation, and damage to their professional lives, including their finances, education and employment.

This intimate image abuse, and the harm it causes, is disproportionately experienced by women. But there is growing trend in financially motivated sexual extortion ('sextortion') in which men are most often the targets.

Service type risk factors:

Research indicates that intimate image abuse occurs particularly on **user-to-user pornography services,** with findings indicating that these services may be at a higher risk of being used by perpetrators to commit the offence. Studies show that intimate image abuse also occurs frequently on **social media services, file-storage and file-sharing services, and dating services**.
Discussion forums and chatrooms have been found to allow its users to form online communities where members are able to discuss and share intimate images, including deepfakes. While **messaging services** are often used by offenders to non-consensually share intimate ages, and to threaten victims with sharing their intimate images.

User base risk factors:

Intimate image abuse is a **gendered** offence; women are more likely to be depicted in the images, and the person committing the offence is more likely to be a man. There is evidence that **age** can be a risk factor, with higher reported cases of intimate image abuse occurring among women aged 18 to 24. Other user groups that face an increased risk of intimate image abuse include people from minority ethnic and racial backgrounds. Additionally, research indicates that **disability,**

[872] Refuge (2023) 'Intimate image abuse – despite increased reports to the police, charging rates remain low' accessed [20 November 2024].

cultural and linguistic diversity, sexual orientation and low **socio-economic status** can increase risk of and from intimate image abuse.

Functionalities and recommender systems risk factors:

Intimate image abuse is primarily committed by **posting images and videos** where perpetrators share intimate images non-consensually.

This risk can be exacerbated by functionalities such as **re-posting and forwarding content, direct messaging and group messaging**, where intimate images can be further shared with larger audiences. Some services allow users to **screen capture, record, or download content**, such as intimate images, which can then be shared on other services.

Several additional functionalities on user-to-user (U2U) services can facilitate the non-consensual sharing of intimate images. **Encrypted messaging** can be used to share intimate images and more easily evade detection.

Livestreaming can be used in the commission of intimate image abuse, where videos of people engaging in sexual activities are broadcast online without their consent. **User groups** allow like-minded individuals to form communities and potentially share intimate image abuse content with one another.

The ability to label or **tag content** can facilitate intimate image abuse, as these tags can be manipulated by users to ensure that intimate images are shown to users who are more likely to know the person depicted. Perpetrators can create **fake user profiles** that impersonate their targets as well as **anonymous user profiles** that gives them added confidence in sharing intimate images without being identified.

With the advent of Generative AI, functionalities that allow for the **editing of visual media** has become central to the creation of deepfake intimate images. These functionalities can vary from purpose-built websites to apps, or bots on U2U services.

Introduction

6.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the intimate image abuse offences listed under 'Relevant offences' in this chapter; and
- The use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

6.2 We set out the characteristics of U2U services that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the effect of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind. Where appropriate, we also consider the

wider *societal* effects of the harm caused to individuals because of exposure to the content or activities that amount to relevant offences. In this case, this would relate to the normalisation of intimate image abuse in some contexts.

6.3 Throughout our risk assessment, we will use the term 'intimate image abuse' to refer to the offences covered in this chapter.[873] This chapter will cover adult intimate abuse; intimate image abuse relating to under-18s is covered in the chapter 'Child sexual exploitation and abuse (CSEA)'.

Relevant offences

6.4 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding intimate image abuse, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

6.5 The priority offences for intimate image abuse are the following:[874]

- A base offence of sharing an intimate image without consent and two more serious offences based on intent to cause humiliation, alarm or distress, and for obtaining sexual gratification. [875]
- Threatening to share an intimate image.

6.6 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and in relation to offences in Scotland, being involved art and part in the commission of these offences).

6.7 The legislation describes the relevant offences by reference to photographs or film that show a person "in an intimate state". However, for the purpose of this chapter we refer to the more commonly used term, 'intimate image'. Most commonly, an 'intimate image' is a photograph or video where the person or people are depicted engaging or participating or are present during a sexual act and/or where their genitals, buttocks or breasts are exposed or covered only with underwear.[876] Intimate image abuse occurs when these intimate images are shared or distributed without the consent of the person pictured; or when someone threatens to share or distribute these images or videos without consent. Intimate image abuse is a gendered harm that disproportionately impacts women, with women around five times more likely to be victims of intimate images abuse than other genders.[877]

---

[873] Intimate image abuse can also be referred to as 'revenge porn' or 'image-based sexual abuse'. 'Revenge porn' is a commonly-used term, but does not adequately capture the power dynamics and the type of content involved in intimate image abuse and we do not therefore use it in our risk assessment. 'Image-based sexual abuse' is often used to describe a broader range of harms than those covered in this chapter, including offences such as cyberflashing and the production or sharing of child sexual abuse material. We will therefore not be using the term 'image-based sexual abuse' in this chapter. These offences are covered separately in the chapters 'Cyberflashing' and 'Child sexual exploitation and abuse (CSEA)'.

[874] Manipulated images and videos, such as deepfakes, are considered within the scope of this offence. Any photograph or video which appears to depict an intimate situation should be treated as a photograph or video actually depicting such a situation. For more detail, refer to the refer to the Illegal Content Judgements Guidance (ICJG).

[875] Section 66(B) of the Sexual Offences Act 2003; section 2 of the Abusive Behaviour and Sexual Harm (Scotland) Act 2016 (asp 22).

[876] We are aware that interpretations of what is 'intimate' can vary among different cultural and religious groups, but Ofcom's interpretation is based on the OSA legislation. The definition of intimate state in the Act is broad and can include "*doing a thing that a reasonable person would consider to be sexual*"; "*the person in an act of urination or defecation*"; and "*the person carrying out an act of personal care associated with the person's urination, defecation or genital or anal discharge*". The majority of our evidence focuses on sexual images.

[877] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 3 August 2023].

Intimate image abuse can form part of a wider continuum of online and offline behaviours by a partner or former partner, and often exhibits the gender dynamics of partner abuse/domestic abuse.[878] These are closely linked to other behaviours explored in greater detail in the Threats, Harassment and Stalking and Coercive Controlling Behaviour chapters.

6.8 Many intimate images are taken, made or shared consensually, and then shared subsequently without consent. However, some intimate image abuse can involve images taken or made without consent. This includes images produced through 'upskirting'[879] and 'downblousing'[880], intimate images made as deepfakes (see next paragraph) or images taken via hacking or using hidden cameras. It also includes intimate images produced using screen capture technology to create a permanent record of a livestream or ephemeral message. The taking and/or making of intimate images is not covered in this chapter. However, once these intimate images are brought onto U2U services, through threatening to share or sharing the images that were made or taken non-consensually, they are within scope of this chapter.

6.9 Deepfakes are forms of audio-visual content that have been generated or manipulated using AI, that misrepresent someone or something.[881] They can involve the creation and/or modification of videos, images or audio including to create realistic sexual synthetic content, such as the alteration of an existing image to 'nudify' the subject, superimposing the face of a person onto the body of another person in a video or image, or the creation of entirely new synthetic content, such as sexual synthetic content of existing individuals. Deepfakes are shared as user generated content on U2U services but could also be created using functionalities present on U2U services.

6.10 For more details on the offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

How intimate image abuse offences manifest online

6.11 This section is an overview which looks at how intimate image abuse manifests online, and how individuals may be at risk of harm. Intimate image abuse covers threats to share intimate images and the non-consensual sharing of intimate images, including deepfake intimate images.

Threats to share intimate images

6.12 Threats to share intimate images predominantly fall into two categories, although they are not mutually exclusive:[882]

---

[878] Henry, N., Flyn, A., Powell, A. 2019. Responding to 'revenge pornography': Prevalence, nature and impacts. [accessed 08 November 2024].

[879] Upskirting refers to someone taking a picture under another person's clothing without their knowledge, with the intention of viewing their genitals or buttocks (with or without underwear). Source: Ministry of Justice, 2019. Upskirting: know your rights. [accessed 3 August 2023].

[880] 'Downblousing' refers to someone taking a photo down a woman's top without consent. Source: Ministry of Justice, 2022. New laws to better protect victims from abuse of intimate images. [accessed 3 August 2023].

[881] Ofcom, 2024, Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes, accessed [18 August 2024].

[882] Although the majority of threats to share pertain to these two main groupings, threats to share are made for other reasons. For example, this may include threats to stop victims and survivors of sexual abuse from disclosing the abuse or

a) as part of a wider pattern of domestic abuse or coercive and controlling behaviour;
b) or for financial or other gain, including sexual extortion (financially motivated sexual extortion, or 'sextortion').[883]

6.13 As with the sharing of intimate images, threats to share intimate images can form part of domestic abuse when current or former partners threaten to share intimate images to coerce or exert control over their partner or former partner. Refuge's 2020 survey found that for 23% of women who had experienced threats to share intimate images, the image was subsequently shared.[884]This behaviour is often part of a pattern of wider abusive behaviour both offline and online and can be part of attempts to make financial or other gains from victim-survivors.[885] See also Harassment, Stalking and Threat Offences (p. 121) and Controlling and coercive behaviour chapters (p. 143) to understand the abuse dynamics between partners or former partners that underpin intimate image abuse.

6.14 In some cases, threats to share are used to blackmail victims and survivors into performing sexual acts or sending more intimate images, and in other cases sending money.[886]

6.15 Threats to share intimate images can also often form part of criminal exploitation, through financially motivated sexual extortion – often referred to as 'sextortion'[887] – or by gangs threatening to share intimate images to force individuals to take part in criminal activity.[888] For example, almost 400 people in Lincolnshire were victims of a recent sextortion scam costing nearly £60k in total between April 2023 and March 2024, where 45% of reports received by police involved use of Instagram or Snapchat, as well as Facebook, Whatsapp and dating sites.[889] Though sextortion can happen to individuals of any age and gender, 93% of sextortion cases dealt with by the Revenge Porn Helpline[890] in 2023 were reported by men, with the Helpline seeing a 54% rise in sextortion cases compared to 2022.[891] Additionally, the IWF in the first half of 2024 has noted a 19% rise in sextortion cases

---

threats to "out" an individual's sexuality or other aspects of their life. Source: Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

[883] Revenge Porn Helpline, 2024. Revenge Porn Helpline 2023 Report. [accessed 10 October 2024].

[884] Refuge, 2020. The Naked Threat. [accessed 3 August 2023].

[885] Refuge found that 83% of women who had experienced threats to share from their partners or ex-partners had also experienced other forms of abuse. Other types of abuse experienced also included emotional abuse (43%), coercive and controlling behaviour (39%), sexual abuse (26%), other tech abuse (20%), physical abuse (17%) and economic abuse (15%). Source: Refuge, 2020.

[886] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023]

[887] This can take many different forms. For example, some perpetrators will use fake user profiles on dating services to connect with individuals and extract intimate images from them, which they will then threaten to disclose unless money is sent. 'Sextortion', as it is also referred to, is also used by some to specifically describe 'webcam blackmail', which involves criminals befriending individuals online and persuading them to perform sexual acts over camera. These are captured or recorded by the perpetrator, who then blackmails the victim and survivor by threatening to share the images if they do not send money. Source: Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023] ; Metropolitan Police, no date. Sextortion. [accessed 3 August 2023].

[888] Storrod, M. and Densley, J. 2017. 'Going viral' and 'Going country': the expressive and instrumental activities of street gangs on social media; Harvard, T. E., Densley, J. A., Whittaker, A. and Wills, J., 2021. Street gangs and coercive control: The gendered exploitation of young women and girls in county lines, *Criminology & Criminal Justice,* 23(3), 313-329.

[889] Lincoldnshire Police. 2024. Warning over rise in 'sextortion' cases'. [accessed 10 October 2024].

[890] Revenge Porn Helpline is a UK-based organisation which supports adult victims of intimate image abuse.

[891] Revenge Porn Helpline, 2024. Revenge Porn Helpline 2023 Report. [accessed 10 October 2024].

involving children compared to the same period in 2023, particularly those involving teenage boys.[892] For further information see also the Register of Risks chapter 'CSEA'.

6.16 To put the risks of this offence into context, one in 14 adults in England and Wales have experienced a threat to share intimate images.[893] There are indications that instances of threats to share intimate images are rising in the UK. Total reports, including non-consensual sharing of intimate images, to the Revenge Porn Helpline rose from 521 in 2015, to 18,426 in 2024, with threats to share making up 8% of this number.[894]

Non-consensual sharing of intimate images

6.17 Between 2015 and 2021, over 28,000 reports of the disclosure of sexual images without consent were recorded by police.[895] The number of intimate image abuse offences recorded increased 40% between 2020 and 2021.[896]

6.18 Non-consensual sharing can be perpetrated in the following ways, noting that these are not mutually exclusive, nor is this list comprehensive:

- A user shares an image as part of a pattern of other abusive or controlling behaviours;
- A user or users 'collect', 'share' or 'trade' intimate images as part of on and offline misogynistic group dynamics, where social status, humour and sexual gratification are motivating factors ('collector culture');
- A user shares an image in order to seek financial gain, including 'sextortion';[897]
- Users re-share or re-post already uploaded non-consensual intimate images, causing secondary abuse or re-victimisation as the images continue to circulate.

6.19 In the context of domestic abuse, a partner or former partner may share intimate images of victims and survivors without consent to exert control or coercion. See also the chapters on 'Harassment, stalking, threats and abuse' and 'Controlling and coercive behaviour' to understand the abuse dynamics between partners or former partners that underpin intimate image abuse.

6.20 Intimate image abuse can also take the form of 'collector culture'. This refers to groups of users who, often anonymously, procure, exchange and discuss intimate images of women without their consent.[898] Some groups name victims and survivors, predominantly women, and share personal information such as name, address, family members and social media profiles (also known as 'doxing').[899] There are various websites that support these and similar communities of perpetrators, ranging from public to private across adult sites,

---

[892] Internet Watch Foundation. 2024. 'Teenage boys targeted as hotline sees 'heartbreaking' increase in child 'sextortion' reports'. [accessed 10 October 2024]. Note: Intimate Image abuse offences involving those under the age of 18 are covered by CSEA offences [LINK CHAPTER].
[893] Ministry of Justice, 2022. New laws to better protect victims from abuse of intimate images. [accessed 4 August 2023].
[894] Revenge Porn Helpline, 2024. Revenge Porn Helpline 2023 Report. [accessed 10 October 2024].
[895] Ministry of Justice, 2022. New laws to better protect victims from abuse of intimate images. [accessed 3 August 2023].
[896] Refuge, 2023. Intimate image abuse – despite increased reports to the police, charging rates remained low. [accessed 31 August 2023].
[897] Sextortion is a form of blackmail that involves threatening to publish sexual information, photos or videos about someone. Source: Metropolitan Police, n.d. Sextortion. [accessed 4 August 2023].
[898] Moore, A., 2022. 'I have moments of shame I can't control': the lives ruined by explicit 'collector culture', The Guardian, 6 January. [accessed 3 August 2023].
[899] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 3 August 2023].

image-boards, community forums and specific 'revenge sites' and varying in theme as well as in functionality.[900]

6.21 Regarding the sharing of intimate images to seek financial gain – financially motivated sexual extortion, sometimes known as 'sextortion' – recent data from the Revenge Porn Helpline showed that in one in five cases (22%), victims confirmed that their intimate content had been shared. But in half of cases (54%) they were unsure whether private content had been made public or circulated online.[901] This is notable increase from 2021 when significantly fewer cases involved victims reporting their intimate content had been shared.[902]

6.22 In addition, once images are online on a single site, they can proliferate; in Revenge Porn Helpline data between 2018 and 2022, 35 cases alone generated 16,937 images shared across 327 platforms, averaging 484 images per victim.[903] This is a consequence of images being shared numerous times on platforms, and across different platforms. This can happen through manual resharing of content and through the design of the service itself; for example, smaller adult sites will use web-crawling software to scrape content from larger, more established adult sites. Academics have identified such software on adult sites as playing a vital role in non-consensual intimate image distribution.[904]

Deepfake intimate image abuse

6.23 A growing form of intimate image abuse involves the creation and sharing of deepfake intimate images.[905] In 2023, more deepfake abuse videos were posted online than in every previous year combined. On the top 40 websites dedicated to deepfake abuse there were over 270,000 videos, which had gained over 4 billion views, and Google Search was driving 68% of traffic to these sites.[906] The same analysis, from My Image My Choice, also captured a vast increase in apps and services that 'nudify' the target, or place them into a sexual context, with 80% (4 in 5) of the services available launched in the previous 12 months. In January 2024 these apps and services had 40 million visitors.[907]

6.24 In a 2024 Ofcom survey, 14% of adult internet users who believed they had seen a deepfake online in the last 6 months said it was sexual. Of these respondents, 64% (more than 3 in 5) said it was of a celebrity or public figure; 42% (more than 2 in 5) said it was a stranger, and 15% said it was of someone they knew; 6% of these respondents said the deepfake they saw was of themselves.[908] A survey by the Alan Turing Institute and Oxford Internet

---

[900] Henry, N., Flynn, A. 2019. Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support. *Violence against women 25(16)*. [accessed 10 October 2024].
[901] UK Safer Internet Centre, 2024. Sextortion Report – August 2022 to August 2024. [accessed 18 November 2024].
[902] Revenge Porn helpline, 2021. RPH cases and trends of 2021. [accessed 09 October 2024].
[903] Revenge Porn Helpline (Huber, A. and Ward, Z.), 2024. Non-Consensual Intimate Image Distribution: Nature, Removal, and Implications for the Online Safety Act. [accessed 10 October 2024]
[904] Henry, N., Flynn, A. 2019. Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support. *Violence against women 25(16)*. [accessed 10 October 2024].
[905] Flynn, A., Powell, A., Scott, A. J., and Cama, E. (2022). Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse. *The British Journal of Criminology*, 62(6). [accessed 06 February 2024].
[906] My Image My Choice, 2024. Deepfake Abuse: Landscape Analysis 2023-24. [accessed 14 August 2024]
[907] My Image My Choice, 2024.
[908] Ofcom, 2024, Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes, accessed [18 August 24]

Institute found that 19% (nearly 1 in 5) of adult respondents who had encountered a deepfake online said it was a non-consensual sexual deepfake.[909]

6.25 The victims and survivors of deepfake abuse are overwhelmingly women. In 2019, Deeptrace reported a 100% increase in the total volume of deepfake videos online over the previous year, with 96% of these videos consisting of non-consensual sexual deepfakes of women.[910] In the following year, Sensity AI found 'stripped' images of 104,582 women on the messaging service, Telegram, where users utilised an AI 'bot' to create synthetic intimate images, with 70% of targets being private individuals whose photos were taken from social media or private material.[911] A 2022 international survey found that 10% of women and girls aged 16 to 64 had experienced at least one form of deepfake abuse.[912] Many of the apps and websites used to 'nudify' or undress victims from existing images only include functionality to edit women's bodies.[913] [914]

Risks of harm to individuals presented by intimate image abuse

6.26 Many survivors and victims of intimate image abuse experience this as a form of sexual intrusion, with impacts across their everyday lives, relationships, and professional worlds.[915] A meta-analysis of research examining the effects of intimate image abuse found that most quantitative studies recorded a significant association between prior victimisation and depression; some articles found associations between IIA and poor health outcomes, including suicidality and self-harm.[916] 60% of Revenge Porn Helpline clients are referred to a mental health service due to the significant impact of their experiences with IIA.[917]

6.27 Beyond psychological effects, the impact of IIA is felt as a "social rupture [...] a significant devastation" that radically disrupts victim-survivors' lives. Survivors and victims face a constant threat, characterised by the persistence of material that isn't removed, or might resurface. They can develop hypervigilance, become isolated, and face constrained opportunities as a consequence.[918] In addition, many survivors and victims report feeling 'extremely' fearful for their safety because of experiencing IIA, with women more likely than men to report fear for safety.[919]

---

[909] Sippy, T., Enock, F. E., Bright, J. and Margetts, H. Z. 2024 Behind the Deepfake: 8% Create; 90% Concerned: Surveying public exposure to and perceptions of deepfakes in the UK. [accessed 18 August 2024].
[910] Deeptrace, 2019. The State of Deepfakes: Landscape, threats and impact. [accessed 08 November 2024]
[911] Sensity AI, 2020. Automating Image Abuse: Deepfake bots on Telegram. accessed [15 February 2024].
[912] Indecent images of children (under 18s) would be Child Sexual Abuse Material (CSAM), and are covered in the chapter Child Sexual Abuse and Exploitation (CSEA)
[913] My Image My Choice (2024) Deepfake Landscape Analysis 2023-2024. [accessed 20 September 2024].
[914] Where 'nudify' services allow users to share generated content with other users of the service, the nudify service could be considered to be a regulated user-to-user service and therefore the provider of the service would need to comply with the Part 3 duties, including the Illegal Harms duties of the Act. A GenAI tool that is made available by a service provider with the intention of allowing users to generate pornographic content may be in scope of the Part 5 duties.
[915] McGlynn, C., Johnson, K., Rackley, E., Henry, N., Gavey, N., Flynn, A., and Powell, A. 2021. ''It's Torture for the Soul': The Harms of Image-Based Sexual Abuse. *Social & Legal Studies*, 30(4). [accessed 1 February 2024].
[916] Patel, U. and Roesch, R. 2020. The Prevalence of Technology-Facilitated Sexual Violence: A Meta-Analysis and Systematic Review. *Trauma, Violence and Abuse* 23(2). [accessed 08 November 2024].
[917] Revenge Porn Helpline, 2024. Revenge Porn Helpline 2023 Report. [accessed 10 October 2024].
[918] McGlynn, C., Johnson, K., Rackley, E., Henry, N., Gavey, N., Flynn, A., and Powell, A. 2021.
[919] Henry, N., Flynn, A. and Powell, A. 2019. Responding to 'revenge pornography': Prevalence, nature and impacts. [accessed 08 November 2024]

6.28 Threats to share intimate images can go on for a significant amount of time, and victims and survivors will often receive multiple threats throughout the day. This is particularly common where threats form part of a wider pattern of domestic abuse, which may go on for months or years. Regular threats are also often experienced by victims and survivors of threats for financial or other gain.[920]

6.29 Such threats can also affect other areas of survivors and victims' lives. A study by Thorn[921] looking at the experiences of young people who had been targets of threats to expose sexual images[922], found 41% (more than 2 in 5) of respondents had lost a relationship with a friend or family member or partner because of the incident. Twelve percent moved home, 10% (1 in 10) of respondents reported problems at school and 8% reported problems in their jobs.[923] Similarly, Refuge found that threats to share intimate images as part of domestic abuse had resulted in some victims and survivors allowing the perpetrator to have contact with their children, continuing or resuming their relationship with the perpetrators, or telling the perpetrator where they currently lived.[924]

6.30 It is important to note that those who experience deepfake intimate image abuse can face equally severe and life-changing consequences as a result. Flynn et al, across 75 interviews with survivors and victims and stakeholders, identify a range of serious and ongoing harms, including emotional, psychological, professional and relational effects, many occurring well beyond the occurrence of the initial abuse.[925]

6.31 Research has found that men are the primary perpetrators of intimate image abuse. A study by Cyber Civil Rights Initiative[926] found that men were twice as likely as women to report being the perpetrators of intimate image abuse.[927] Where the gender of the perpetrator was known, twice as many men than women were the perpetrators of threats.[928]

6.32 Survivors and victims may struggle with a range of feelings on the disclosure of intimate images, including shame, helplessness, self-blame, isolation and humiliation.[929] It can affect an individual's feeling of security, and damage their professional lives, including their employment, education and careers, with a knock-on financial effect.[930]

---

[920] A survey by Thorn, looking at the experiences of young people who had been targets of threats to expose sexual images, found that 34% of respondents had received threats on a daily basis. In the same survey, 22% of respondents reported the threats lasting for more than six months. (Sample was recruited online and consisted of 1,631 18 to 25 year olds. Facebook and Twitter provided Thorn with grants which were used to fund the recruitment of respondents). Source: Thorn (Wolak, J. and Finkelhor, D.), 2016. Sextortion: Findings from a survey of 1,631 victims. [accessed 4 August 2023].

[921] Thorn is an international non-profit organisation working to develop new technologies to combat online child sexual abuse.

[922] Sample was recruited online and consisted of 1,631 18- to 25-year-olds who had been targets of threats to expose sexual images. Facebook and Twitter provided Thorn with grants which were used to fund the recruitment of respondents.

[923] Thorn (Wolak, J. and Finkelhor, D.), 2016. Sextortion: Findings from a survey of 1,631 victims. [accessed 4 August 2023].

[924] Refuge, 2020. The Naked Threat. [accessed 4 August 2023].

[925] Flynn, A., Powell, A., Scott, A. J., and Cama, E. (2022). Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse. *The British Journal of Criminology*, 62(6). [accessed 06 February 2024].

[926] The Cyber Civil Rights Initiative is a non-profit organisation based in the United States working to combat online abuses that threaten civil rights and civil liberties.

[927] Cyber Civil Rights Initiative (Eaton, A. A., and Jacobs, H. and Ruvalcaba, Y.), 2017. 2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration: A Summary Report. [accessed 4 August 2023].

[928] Refuge, 2020. The Naked Threat. [accessed 4 August 2023].

[929] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

[930] Davidson, J., Livingstone, S., Jenkins, S., Gekoski, A., Choak, C., Ike, T. and Phillips, K., 2019. Adult Online Hate, Harassment and Abuse: A rapid evidence assessment. [accessed 4 August 2023].

6.33 All gendered forms of online abuse, including intimate image abuse, have a 'chilling' effect on women and girls' freedom of speech. Intimate image abuse, including hacking and deepfake creation, often targets women in public life, such as journalists and politicians.[931] A 2022 survey of international women journalists found that 15% had experienced image-based abuse, such as manipulation of photos and videos, stolen images, and the non-consensual sharing of intimate images. 29 respondents reported deepfakes and other forms of synthetic media attacks.[932] However, it is important to note that intimate image abuse of all forms predominantly affects women without such public presence.

6.34 For women in the public eye, such as politicians and journalists[933], and for other individuals, the experience or fear of intimate image abuse can force them to constrain their online behaviours and expression. Fifty percent of women fear being the target of image-based sexual abuse compared to 30% (nearly one third) of men.[934] At a societal level this results in the silencing and exclusion of women and girls from online spaces.

6.35 The impact of intimate image abuse can vary substantially based on an individual's personal circumstances and the cultural or social context. This can lead to different risks of harm, including risking stigma within their community which expands out to their family members.[935] These risks can align with different understandings of what 'intimate' might mean.[936]

Evidence of risk factors on user-to-user services

6.36 We consider that the risk factors below are liable to increase the risks of harm relating to intimate image abuse offences. This is also summarised at the start of the chapter.

Risk factors: Service type

6.37 Research indicates that the following types of services are used to facilitate or commit offences related to intimate image abuse: user-to-user pornography services, social media services, messaging services, file-storage and file-sharing services, dating services, and video-sharing services.

6.38 An analysis of seventy-seven services hosting intimate image abuse found content likely to be intimate image abuse across six types of services: dedicated intimate-image abuse sites,

---

[931] BBC News, 2016, Celebgate hack: Man to plead guilty to nude photos hack, BBC News, 15 March. [accessed 3 August 2023]; Channel 4 News, 2024, Exclusive: Hundreds of British celebrities victims of deepfake porn, Channel 4 News, 21 March [accessed 3 September 2024].

[932] The sample comprised of 714 "women-identifying" journalists. Source: Posetti, J. & Shabbir, N. 2023. The Chilling: A global study of online violence against women journalists. [accessed 3 September 2024].

[933] When intimate image abuse is leveraged against women in the public eye, often alongside other forms of abuse, they face reputational damage. 10% of women journalists surveyed in 2022 said their professional reputations or employment had been affected as a result of online abuse. Source: Posetti, J. & Shabbir, N. 2023. The Chilling: A global study of online violence against women journalists. [accessed 3 September 2024].

[934] Enock, F. E., Stevens, F., Bright, J., Cross, M., Johansson, P., Wajcman, J., Margetts, H. Z. 2024. Understanding gender differences in experiences and concerns surrounding online harms: A short report on a nationally representative survey of UK adults. *Computers and Society* (forthcoming). [accessed 14 November 2024].

[935] Refuge, 2022. Marked as Unsafe: How online platforms are failing domestic abuse survivors. [accessed 22 August 2023].

[936] See, for example, the discussion of a broader idea of 'intimate image' in Rackley, E., McGlynn, C., Johnson, K., Henry, N., Gavey, N, Flynn, A. and Powell, A. 2021. Seeking justice and redress for victim-survivors of image-based sexual abuse. *Feminist Legal Studies,* Volume 29. [accessed 14 November 2024].

user-generated pornography sites[937], image boards, community forums, blogging platforms, and social media services.[938] In a random sample of Revenge Porn Helpline data from between 2018 and 2022, 52% of the 200 cases involved distribution to social media, with Facebook and Instagram being involved in the greatest number of cases, followed by Twitter and Snapchat. Distribution via public URLs was linked to 44% (more than 2 in 5) of cases.[939]

User-to-user pornography services

6.39 User-to-user pornography services, in particular, host significant amounts of intimate image abuse. The Revenge Porn Helpline sets out that services which offer pornographic content may be at higher risk of being used by perpetrators to engage in intimate image abuse.[940] A study of mainstream pornography sites found that, of the analysable homepage videos, 2.2% had titles which constituted descriptions of intimate image abuse.[941] Another study found that a user-to-user pornography website hosted nearly 60,000 videos under four phrases associated with intimate image abuse.[942] Additionally, of the intimate images that were reported by the Revenge Porn Hotline in 2023, 28% of the content was found on user-to-user pornography websites.[943]

6.40 A study for Australia's eSafety Commissioner, which analysed content on a "*user-generated porn site with an online community of those who want to view and share non-consensual, amateur images"* found that a search for 'revenge' on the site came up with over 12,000 images and 11,000 videos – all material that was likely to have been shared without consent.[944]

Social media services

6.41 There is also evidence of a significant amount of intimate image abuse taking place on social media services. The Revenge Porn Helpline found that Instagram, Snapchat, and X (formerly Twitter) accounted for 17%, 6% and 4%, respectively, of reported cases of intimate image abuse in 2023.[945] A case reported by Refuge described a perpetrator hacking into their ex-partner's Instagram account, locking her out of it, making the account public and then uploading intimate images of her.[946] In some cases, intimate images are shared on social media services, and then re-shared to other sites.[947]

---

[937] While the analysis only applies to online user-to-user pornography services that allow users to share user-generated pornographic content, it is possible that some of our evidence also covers services that allow users to view pornographic content that has been produced by providers of pornographic content.
[938] Henry, N. and Flynn, A., 2019. Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support, *Violence Against Women,* 25(16), pp.1932-1955. [accessed 5 September 2023].
[939] Revenge Porn Helpline (Huber, A. and Ward, Z.), 2024. Non-Consensual Intimate Image Distribution: Nature, Removal, and Implications for the Online Safety Act. [accessed 10 October 2024].
[940] Revenge Porn Helpline's response to 2021 Law Commission consultation. [accessed 3 August 2023].
[941] Vera-Gray, F., McGlynn, C., Kureshi, I. and Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology,* 61(5). [accessed 22 August 2023].
[942] Henry, N. and Flynn, A., 2019. Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support, *Violence Against Women,* 25(16), pp.1932-1955. [accessed 5 September 2023].
[943] Revenge Porn Helpline, 2024. Revenge Porn Helpline 2023 Report. [accessed 10 October 2024].
[944] Australian Government (Office of the eSafety Commissioner), 2017. Image-based abuse. National survey: summary report. [accessed 22 August 2023].
[945] Revenge Porn Helpline, 2024.
[946] Refuge, 2022. Marked as Unsafe: How online platforms are failing domestic abuse survivors. [accessed 22 August 2023].
[947] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

Messaging services

6.42 Private messaging services, and the direct messaging functionality in particular, are often used for both the non-consensual sharing of intimate images offence and the threat to share intimate images offence. The Revenge Porn Helpline highlighted private messages as a source of reported cases of intimate image abuse, representing 17% of cases where images were shared in 2023. This includes private messaging services as well as emails and texts.[948] A Thorn survey[949] looking at the experiences of young people who had been targets of threats to expose sexual images found for 41% of respondents the perpetrators had used direct messaging services to contact them, including the sending of threats to share intimate images.[950]

File-storage and file-sharing services

6.43 There is some evidence to suggest that file-storage and file-sharing services are a risk factor, and there are reported cases of file-sharing services being used to host or share intimate images.[951] In 2019, Police Scotland investigated a file-sharing service after finding a series of folders and subfolders of intimate images of women. The folders were organised by regions and cities across the UK, then by the names of the women.[952] Recent analysis of case data from the Revenge Porn Helpline identified file-sharing services as a location for the collection of and sharing of intimate images, where large collections of images can be made and shared with a single link; finding that file sharing sites made up 1% of the sample of public URLs from cases between 2018 and 2022. In one case, 940 images were found in a collection on a file-sharing site and were able to be shared as a single file. File-sharing is linked with discussion forums and chatrooms, where chat threads will use file-sharing links to share large amounts of IIA content in forums and threads.[953]

Dating services

6.44 There is some evidence to suggest that dating services are often used to perpetrate intimate image abuse offences, particularly in relation to sextortion. The Revenge Porn Helpline reported that in cases of sextortion, the perpetrators often set up fake profiles on online dating services which they use to extract intimate images of the victims and survivors.[954] A survey by Thorn of young people who had been targets of threats to expose sexual images found that in 11% of cases the perpetrators and the victims and survivors first interacted through a dating service.[955]

---

[948] Revenge Porn Helpline, 2024.
[949] Sample was recruited online and consisted of 1,631 18- to 25-year-olds who had been targets of threats to expose sexual images. Facebook and Twitter provided Thorn with grants which were used to fund the recruitment of respondents.
[950] Thorn (Wolak, J. and Finkelhor, D.), 2016. Sextortion: Findings from a survey of 1,631 victims. [accessed 4 August 2023].
[951] McLaughlin, E., 2018. Dropbox removed folder containing explicit photos of female service members, ABC News, 12 March. [accessed 22 August 2023].
[952] BBC News, 2019. Victim's warning after finding revenge porn from 'every UK city', BBC News, 17 May. [accessed 22 August 2023].
[953] Revenge Porn Helpline (Huber, A. and Ward, Z.), 2024. Non-Consensual Intimate Image Distribution: Nature, Removal, and Implications for the Online Safety Act. [accessed 10 October 2024].
[954] Revenge Porn Helpline, n.d. What to do if you've been victim to online webcam blackmail, also known as sextortion. [accessed 12 September].
[955] Sample was recruited online and consisted of 1,631 18-25-year-olds who had been targets of threats to expose sexual images. Source: Thorn (Wolak, J. and Finkelhor, D.), 2016. Sextortion: Findings from a survey of 1,631 victims. [accessed 12 September 2023].

Video-sharing services

6.45 There is some evidence to suggest that video-sharing services, particularly those with livestreaming, are used to commit intimate image abuse.[956] Thorn found that for 23% of respondents, perpetrators contacted victims and survivors through video-sharing services.[957]

Discussion forums and chatrooms

6.46 There is evidence to suggest that discussion forums and chatrooms are often used to perpetrate intimate image abuse offences. Henry and Flynn's 2019 study identified community forums that allowed members to form communities and 'sub-forums' where they discussed and shared intimate images, including deepfakes. Members were able to share non-consensual intimate images via 'threads' that discuss specific topics.[958] In the Revenge Porn Helpline's analysis of cases between 2018 and 2022, 18% of URLs where intimate images were found were forums for sexual content and chat threads which allow users to share sexual content anonymously. These sites can house large collections and often contain personal details of the victims.[959]

6.47 A significant amount of deepfake content is hosted on dedicated sites, with many 'creators' congregating on online message boards or forums of dedicated websites.[960] Conversations in these forums can involve, writing verbal abuse about their victims, or 'doxxing' victims by broadcasting personal information about them.[961] Often, shared images are categorised geographically, meaning those depicted could be located offline.[962] Sites have also been known to run designated U2U marketplaces where creators advertise customised deepfakes in exchange for a fee, with some creators allegedly earning over $20,000 a month.[963] Research group Graphika found a 2000% increase in referral links to these kinds of specific sites from U2U services such as Reddit and X in 2023; as well as finding a million users on Telegram taking part in deepfake abuse services on the site.[964]

Risk factors: User base

User base size

6.48 There is a risk of intimate image abuse-related offences occurring both on services with a large user base and those with and on those with smaller ones, with different reasons for each type.

6.49 Sometimes perpetrators share this content on services with a large user base because more people will see the content there. A study found that highly visible sites such as social

---

[956] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[957] Sample was recruited online and consisted of 1,631 18-25-year-olds who had been targets of threats to expose sexual images. Source: Thorn (Wolak, J. and Finkelhor, D.), 2016.
[958] Henry, N. and Flynn, A., 2019. Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support, *Violence Against Women,* 25(16), pp.1932-1955. [accessed 10 October 2024].
[959] Revenge Porn Helpline (Huber, A. and Ward, Z.,), 2024.
[960] Tenbarge, K. (2023). The Deepfake porn industry is operating in plain sight, NBC News, 27 March. [accessed 14 August 2024]; My Image My Choice, 2024. Deepfake Landscape Analysis 2023-2024, [accessed 20 September 2024].
[961] My Image My Choice, 2024.
[962] My Image My Choice, 2024.
[963] My Image My Choice, 2024.
[964] Graphika, 2023. A Revealing Picture. [accessed 14 August 2024].

media services are considered by perpetrators as a place where their material will be seen by people whom the victims and survivors know. The primary motivation for the perpetrator is shaming their target.[965] For intimate image abuse as a form of domestic abuse, the perpetrator would be looking for the largest number of users known to the person they are abusing.

6.50 Conversely, some perpetrators are drawn to less visible services with smaller user bases. It is possible that on smaller services the images are less likely to be discovered by the victims and survivors, so there is less chance that the material will be reported. Research shows that perpetrators may be drawn to some types of *"smaller and less well-regulated sites"* to share intimate images, particularly those based in countries with less stringent copyright laws. This may be because images are less likely to be taken down.[966]

User base demographics

6.51 The following section outlines key evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

6.52 Data suggests that user base characteristics including gender, age, race and ethnicity, cultural and linguistic diversity, disability, socio-economic status and sexual orientation could lead to an increased risks of harm to individuals.

6.53 Intimate image abuse is a gendered harm. Nearly all research on intimate image abuse finds that women are significantly more likely than men to experience this abuse and are more likely to have had more images shared in the process. The Revenge Porn Hotline found that for the average number of images it reported, where the victim and survivor was a woman, was 8.6 images per victim. For men the average number of images reported was 0.3 images per victim.[967] In a single month of data from August 2023, it reported 708 images for women clients and 34 for men. In addition, in this month, 15 women experienced voyeurism, compared to 4 men.[968]

6.54 Similarly, the Law Commission identified that victims and survivors of intimate image abuse are primarily women, and the perpetrators are primarily men.[969] Revenge Porn Helpline data from 2023 found that where the perpetrator was known, 82% of female clients had a male perpetrator, compared to male clients, for whom 90% of perpetrators were 'criminal gangs'.[970] In Flynn et. al's international survey on deepfake abuse, they found that men were twice as likely as women to self-report engaging (creating, sharing or threatening to share) in some form of deepfake abuse.[971] Research by Revealing Reality about non-consensual sharing of intimate images among young people found that, predominantly,

---

[965] Office of the eSafety Commissioner, 2017. Image-based abuse. National survey: summary report. [accessed 22 August 2023].
[966] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[967] Revenge Porn Helpline, 2024. Revenge Porn Helpline 2023 Report. [accessed 10 October 2024].
[968] Revenge Porn Helpline (Huber, A. and Ward, Z.), 2024. Non-Consensual Intimate Image Distribution: Nature, Removal, and Implications for the Online Safety Act. [accessed 10 October 2024].
[969] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[970] Revenge Porn Helpline (2024) *Revenge Porn Helpline 2023 Report*. [accessed 10 October 2024].
[971] Flynn, A., Powell, A., Scott, A. J., and Cama, E. (2022). Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse. *The British Journal of Criminology*, 62(6). [accessed 06 February 2024]. Note: The creation of deepfake intimate imagery is not currently an offence under the OSA.

nudes were shared non-consensually by men and boys who were in more 'casual' relationships that were short-term or less serious, and the victims were women and girls.[972]

6.55 As outlined previously, many cases of intimate abuse are strongly tied to domestic abuse, with perpetrators trying to control the survivors and victims. Of the 376 prosecutions for intimate image abuse offences in the year ending March 2019, 83% were domestic abuse related.[973] Revenge Porn Helpline data from 2023 found that where the perpetrator was known, 67% of survivors and victims had a perpetrator who was a current or previous partner.[974]

6.56 For the most part, threats to share intimate images are also disproportionately experienced by women. Refuge estimates that one in seven women aged 18 to 34 in England and Wales have experienced threats to share their intimate images, in comparison to one in nine men aged 18 to 34.[975] Similarly, research by the Cyber Civil Rights Initiative found that in the USA, women were two and a half times more likely than men to have received a threat to share an intimate image.[976] The exception to this is sextortion for money, which is disproportionately experienced by men;[977] recent Revenge Porn helpline data finds that 73% of men accessing the service had been victimised by sextortion, compared to 6% of women accessing the service.[978] A recent study of sextortion in the US during the COVID-19 pandemic also found that the victims and survivors were more likely to be men.[979]

6.57 In contrast, threats to share intimate images to coerce an individual into sharing more intimate images is predominantly experienced by women.[980]

6.58 Age is a risk factor in a user's experience of intimate image abuse. Ofcom research has found that intimate image abuse appears to be a harm experienced online to a higher degree by younger adults. According to our research, 7% of internet users aged 18 to 24 had seen or experienced 'sharing of, or threats to share, intimate images without consent' in the four weeks leading up to participating in the research, compared to 3% of all adult internet users.[981] Additionally, 20% of respondents aged 16 to 19 years in Flynn et al.'s 2022 survey compared to 3.4% of respondents aged 50 to 64 years reported experiencing at least one form of deepfake abuse.[982]

6.59 As with many online harms, the risk of intimate image abuse is likely to be higher for people with protected characteristics.

---

[972] Revealing Reality, 2023. Without Consent. [accessed 15 August 2024]
[973] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[974] Revenge Porn Helpline, 2024. Revenge Porn Helpline 2023 Report. [accessed 10 October 2024].
[975] Refuge, 2020. The Naked Threat. [accessed 4 August 2023].
[976] Cyber Civil Rights Initiative (Eaton, A., and Jacobs, H. and Ruvalcaba, Y.), 2017. 2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration: A Summary Report. [accessed 4 August 2023].
[977] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023]; Revenge Porn Helpline (Ward, Z.), 2022. Revenge Porn Helpline Report. [accessed 22 August 2023].
[978] Revenge Porn Helpline (2024), *Revenge Porn Helpline 2023 Report,* accessed [10 October 2024].
[979] Eaton, A. A., Ramjee, D. and Saunders, J. F., 2022. The Relationships between Sextortion during COVID-19 and Pre-pandemic Intimate Partner Violence: A Large Study of Victimization among Diverse U.S Men and Women, *Victims & Offenders,* 18(2). [accessed 22 August 2023].
[980] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[981] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 21 November 2024].
[982] Flynn, A., Powell, A., Scott, A. J., and Cama, E. (2022). Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse. *The British Journal of Criminology*, 62(6). [accessed 06 February 2024].

6.60 Ofcom research has found that adult internet users with minority ethnic backgrounds were more likely than white adult internet users to report seeing or experiencing the '*sharing of, or threats to share, intimate images without consent*' online in the past four weeks (4% versus 1%).[983] A study looking at the 'sextortion' type of intimate image abuse during the COVID-19 pandemic in the US found that Native American, Alaskan Native women and Black women reported sextortion victimisation more often than women of other ethnicities.[984]

6.61 Sexuality can be a risk factor in people's experience of intimate image abuse. The same study into sextortion during COVID-19 in the US found that lesbian participants, bisexual participants, and those who identified as 'other' sexual orientations reported sextortion victimisation during the pandemic more often than gay or heterosexual participants.[985] Similarly, Ofcom research found that more bisexual adult internet users than heterosexual adult users had reported experiencing the 'sharing of, or threats to share, intimate images without consent' online in the past four weeks (5% versus 3%).[986] In Flynn et al.'s survey on deepfake abuse, 27% of LGBT+ respondents compared to 13% of heterosexual respondents were more likely to report victimisation.[987]

6.62 Socio-economic condition has also been linked with increased risk of intimate image abuse.[988] In Ofcom's Online Experiences Tracker, individuals identified as 'Most financially vulnerable' were more likely (8%) than those who were 'Potentially financially vulnerable' (3%) and 'Least financially vulnerable' (3%) to have experienced 'Sharing of, or threats to share, intimate images without consent' in the last four weeks.

6.63 Disability can also be a risk factor. Ofcom research found that adult internet users with any limiting and effecting conditions had experienced the 'sharing of, or threats to share, intimate images without consent' online in the past four weeks leading up to the research more than users with no limiting or effecting conditions (4% versus 3%).[989]

6.64 Individuals with multiple protected characteristics face multiple and unique risks. For example, 4% of women belonging to minority ethnic groups report the 'sharing of, or threats to share, intimate images without consent' compared to 2% of white women in

---

[983] Ofcom, 2023. Q8 (Table 234) in Online Experiences Tracker: Data Tables (Waves 1 and 2). [accessed 4 September 2023].
[984] 7% of Black, Afro-Caribbean or African women and 5% of Native American or Alaskan Native women reported sextortion, while 2.4% of Latinas, 2% of Asian women, and 0.8% of white women reported sextortion. Source: Eaton, A., Ramjee, D. and Saunders, J., 2022. The Relationships between Sextortion during COVID-19 and Pre-pandemic Intimate Partner Violence: A Large Study of Victimization among Diverse U.S Men and Women, *Victims & Offenders,* 18 (2). [accessed 22 August 2023].
[985] 7.1% of lesbian participants, 8.9% of bisexual participants and 6.3% of participants who identified as "other" sexual orientation reported sextortion, compared to 2.1% of gay participants and 2.9% of heterosexual participants. Source: Eaton, A., Ramjee, D. and Saunders, J., 2022.
[986] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 21 November 2024].
[987] Flynn, A., Powell, A., Scott, A. J., and Cama, E. (2022).
[988] For example, in research conducted by Revealing Reality, girls from more disadvantaged backgrounds were more likely to report that nude images they had shared had been shown and/or distributed without their permission; girls with multiple indicators of disadvantage were twice as likely to report that someone they had sent a picture to had sent it on without consent, compared to less disadvantaged girls. This research was predominantly focused on under 18s, therefore, some of the activity or experiences being discussed would be considered child sexual abuse (see the CSEA chapter), but the relationship identified between negative experiences and socio-economic indicators is stark. Source: Revealing Reality, 2022, Not just flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 18 August 2024].
[989] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 21 November 2024].

Ofcom research.[990] In recent research by the Revenge Porn Helpline, they found that 3 per cent of the websites featuring intimate images were sites which focused on individuals from South Asian descent.[991] Additionally, women belonging to these groups may experience increased effects including inability to return to a home country.[992]

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

6.65 Perpetrators can create fake user profiles which facilitate the commission of intimate image abuse. These can be used to impersonate victims and survivors. A study by Australia's eSafety Commissioner[993] found examples of perpetrators of intimate image abuse setting up fictitious social media accounts under the name of their targets, and then spreading sexual photos of them through these fake profiles, as well as spreading lies about them.[994]

6.66 Perpetrators can also set up multiple user profiles from which they can continue to non-consensually share intimate images, even when individual accounts and their associated user profiles are reported and/or blocked.[995]

Anonymous user profiles

6.67 The ability to set up anonymous user profiles appears to facilitate intimate image abuse. McGlynn and Woods' research found that a large proportion of those who upload user-generated porn do so anonymously. Anonymous profiles facilitate the anonymous sharing of intimate images. They also allow perpetrators to share intimate images with less fear of consequences for their actions. Difficulty identifying users is one of the factors impeding police enforcement, and anonymity can make user identification more difficult.[996]

6.68 Perpetrators who share the intimate images for the purposes of sexual gratification or financial gain may abuse the anonymity some services with smaller user bases offer; this makes it less likely the victims and survivors will realise that their images have been distributed and potentially identify the anonymous perpetrator.[997]

---

[990] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 21 November 2024].
[991] These sites are referred to as "Desi" sites, a term used to identify individuals whose descent comes south Asian countries, including India, Pakistan, Bangladesh, Sri Lanka, the Maldives, Nepal, and Bhutan. In the pornography context, the term 'desi porn' has come to be used colloquially as identifying content which explicitly focuses on performers who can be identified as Indian or Pakistani. Source: Revenge Porn Helpline (Huber, A. and Ward, Z.), 2024. Non-Consensual Intimate Image Distribution: Nature, Removal, and Implications for the Online Safety Act. [accessed 10 October 2024].
[992] Revenge Porn Helpline (Huber, A. and Ward, Z.), 2024.
[993] Study based on information provided by women from culturally and linguistically diverse backgrounds who have experienced technology-facilitated abuse (n=29) and 20 stakeholders who provide support services to women.
[994] eSafety Commissioner, 2019. eSafety for Women from Culturally and Linguistically Diverse Backgrounds: Summary Report. [accessed 22 August 2023].
[995] Threats, Harassment and Stalking chapter includes a study from Refuge where one individual tried to block her former partner, only to find over 120 fake accounts created by him over a few weeks to continue harassing her. Source: Refuge, 2022. Marked as Unsafe: How online platforms are failing domestic abuse survivors. [accessed 22 August 2023].
[996] Woods, L. and McGlynn, C., 2022. Pornography platforms, the EU Digital Services Act and Image-based sexual abuse, Media@LSE, 26 January. [accessed 22 August 2023].
[997] eSafety Commissioner, 2017. Image-based abuse. National survey: summary report. [accessed 22 August 2023]; Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].

User networking

User groups

6.69 The ability to create user groups can help like-minded individuals form communities and provide spaces for them to share illegal content or offer advice on engaging in illegal behaviour. Although we do not have direct evidence pointing to perpetrators using these groups, it is particularly likely that perpetrators who share images anonymously, and those engaging in 'collector culture', will use groups to share intimate images. The reasons why users share these images with others may include social status, humour, sexual gratification and misogyny. We therefore expect that, as is also detailed in the Child sexual exploitation and abuse (CSEA) chapter. it is possible that user groups facilitate intimate image abuse as they can be used by preparators to non-consensually share intimate images of others.

User communications

Livestreaming

6.70 Livestreaming has been used to commit intimate image abuse; for example, when sexual activity is broadcast without people's consent. There have been cases in which people were having consensual sex, but it was being livestreamed without the consent of all those present.[998]

Direct messaging

6.71 Direct messaging is a primary functionality that enables the non-consensual sharing of intimate images. The Revenge Porn Helpline found that non-consensual images were being shared via private messaging services; this was the method used in 18% of the cases where images were shared in 2020.[999]

6.72 It is also likely that direct messaging is used by perpetrators to send online threats to disclose intimate images. As detailed in the Harassment, stalking, threats and abuse chapter, direct messaging is used to harass individuals and allows a perpetrator to communicate their threat directly to their target. Thorn found that for 28% of young people who had been targets of threats to expose sexual images, perpetrators had suggested moving their conversation with the victims and survivors to a specific site or app after the initial contact.[1000] Other studies suggest that threatening behaviour can result in incessant messaging from perpetrators.[1001]

Encrypted messaging

6.73 Encrypted messaging is a functionality that can be used by perpetrators of intimate image abuse to avoid content moderation when sharing intimate images through direct messages. The Revenge Porn Helpline found that, particularly in sextortion cases, intimate images

[998] For example, in one case in Australia a perpetrator filmed consensual sexual activity between himself and a victim and survivor and used a livestream to non-consensually share the image with a second perpetrator. Source: BBC News, 2013. Australia cadets online sex case: Two convicted, BBC News, 28 August. [accessed 22 August 2023].
[999] Revenge Porn Helpline (Ward, Z.), 2021.
[1000] Sample was recruited online and consisted of 1,631 18 to 25 year olds who had been targets of threats to expose sexual images. Source: Thorn (Wolak, J. and Finkelhor, D.), 2016. Sextortion: Findings from a survey of 1,631 victims. [accessed 4 August 2023].
[1001] Refuge, 2021. Unsocial Spaces: make online spaces safer for women and girls. [accessed 22 August 2023].

were often sent through messaging services which were end-to-end encrypted and not proactively searched by a moderation team.[1002]

Group messaging

6.74 Group messaging is a functionality used to non-consensually share intimate images by perpetrators on services. A BBC investigation into Telegram found that it was being used by perpetrators to non-consensually share intimate images with large groups of users – tens of thousands of people.[1003] Australia's eSafety Commissioner found that this type of image-based abuse among young people is normally motivated by the preparators seeking the social status of having been able to solicit naked images from their similarly-aged peers.[1004]

Posting content (images, videos) and re-posting and forwarding content

6.75 Posting content, in particular images and videos, is a key functionality in the commission of intimate image abuse. It allows perpetrators to share content, and in some cases intimate images, in an open channel of communication for numerous users to see.

6.76 Perpetrators have also been known to gain unauthorised access to victims' and survivors' accounts and to post intimate images from there. This can result in the victims and survivors having their accounts disabled. It can also result in close contacts of the victims and survivors, such as family and friends, seeing their intimate images, which can lead to relationship challenges and social isolation.

6.77 If intimate images are posted from the hacked business accounts of victims and survivors, this creates additional problems with their employment and reputation, with a potential financial impact.

6.78 The ability to re-share content through re-posting or forwarding aids the commission of intimate image abuse, as images can be shared onwards to other services or individuals. The perpetrator need not be the same user who initially shared the images.

6.79 This secondary distribution of images can cause non-consensually shared intimate images to 'go viral', as it becomes more and more difficult to get images removed when they are repeatedly re-loaded to the same service or shared to other services. There are cases in which intimate images are first shared on social media services, and from there, posted to user-to-user pornography services.

6.80 This secondary content sharing can be done using a service's forwarding functionalities or re-posting functionalities. It can also be done by the user downloading the image to their device (see Downloading content) and then sharing it on another service.[1005] As well as re-sharing intimate images to services, it is possible that the intimate images are also being forwarded on private messaging channels.

[1002] Revenge Porn Helpline (Ward, Z.), 2022. Revenge Porn Helpline Report. [accessed 22 August 2023].
[1003] BBC World Service Disinformation Team, 2022. Why won't Telegram take down my naked photos?, BBC News, 20 February. [accessed 22 August 2023].
[1004] eSafety Commissioner, 2019. Understanding the attitudes and motivations of adults who engage in image-based abuse. [accessed 22 August 2023].
[1005] In one case, provided by the Law Commission, a woman's ex-partner set up a fake Facebook account in her name and uploaded intimate images of her, which were then viewed and copied to user-to-user pornography services. On one website the picture was viewed over 48,000 times. Source: Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

Transactions and offers

6.81 No evidence was found suggesting that transaction and offer functionalities are a risk factor in the facilitation and commission of these offences. However, it is possible that in-service payment functionalities could be used as part of sextortion, where victims and survivors are threatened with the sharing of intimate images unless they send money to the perpetrator. There is little evidence on the mechanisms through which the money is sent to the perpetrators.

Content exploring

Content tagging

6.82 The evidence indicates that the ability to tag content can facilitate the offence of intimate image abuse, as well as exacerbate the impact on victims and survivors. Tagged or labelled content facilitates searches for specific intimate images and allows perpetrators to share intimate images more widely.

6.83 The Law Commission found some sites dedicated to intimate image abuse which organise their content geographically. This enables visitors to the site to look for threads about the areas where they live and try to find images of people they know or have seen.[1006] The Law Commission also found that often, when intimate images are shared, personal information about the victims and survivors is included, which means the images will appear at or near the top of search results relating to them.[1007]

6.84 The ability to label content allows perpetrators to share intimate images more widely, which has the potential to increase the impact on the victims and survivors. McGlynn *et al.* analysed video titles on mainstream user-to-user pornography services and found that labels such as 'voyeur', 'hidden' and 'upskirt' were common; some of these videos are likely to have been intimate image abuse content.[1008] Campaign group #NotYourPorn[1009] has also found an increase in the number of victims and survivors whose intimate images have been tagged or labelled as 'leaked' or 'stolen'.[1010]

Content storage and capture

Downloading content

6.85 Download functionalities facilitate sharing of intimate images, whereby people unknown to the person in the image can download and share the image on other services. As identified in 'Posting content (images, videos)' and 're-posting and forwarding content' under the header 'User Communications', the secondary distribution of images onto services can cause intimate images to go viral and appear on multiple, usually adult, services. Sometimes this secondary distribution of images is done by re-posting or forwarding on the original services. But it is also likely that in some cases, users download the material from the original service, and then upload it to another service. In this manner, the ability to

[1006] Law Commission, 2021.
[1007] Law Commission, 2021.
[1008] Vera-Gray, F., McGlynn, C., Kureshi, I. and Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology,* 61 (5). [accessed 22 August 2023].
[1009] #NotYourPorn is a UK-based movement focused on protecting non-consenting adults, sex workers and under-18s from image-based sexual abuse.
[1010] Dawson, B., 2020. Revenge Porn Is Being Posted Under a Different Name, Vice, 15 December. [accessed 22 August 2023].

download content facilitates intimate image abuse as it allows users to possess intimate images which they will then share further.

Screen capturing and recording

6.86 Services which allow video calling can be used to commit intimate image abuse, particularly when paired with screen recording or capturing functionalities. For example, intimate images can be created non-consensually by recording video calls.[1011] Thorn reported that in some sextortion cases, perpetrators take screen captures or recordings of intimate video calls, which they can then use to threaten victims and survivors unless they meet their demands, which often include financial demands.[1012]

Content editing

Editing visual media

6.87 Editing visual media is central to the creation of deepfake intimate images discussed earlier in this chapter. Tools and functionalities that allow images to be created or edited can be used to make intimate images and videos without consent.[1013] These functionalities can vary from purpose-built websites to apps or bots on U2U services, such as messaging sites, to open-source software that users can download and customise.[1014] These tools have proliferated since their emergence in 2017, with increasing efficiency, accessibility and sophistication of outputs.[1015] Where early applications were restricted to software that swapped women's faces onto pornographic content, these have now expanded to include nudifying apps, cloning voices, lip-synchronisation changes and facial re-enactment, facilitating easy creation of highly realistic content.[1016]

6.88 A high-profile recent case of deepfake abuse against women in public life involved deepfake nude images of musician Taylor Swift being circulated on X, with one image receiving 47 million views before being removed.[1017] However, it is important to recognise that private individuals are just as vulnerable to this form of abuse, as noted in prior paragraphs.

Recommender systems

6.89 No evidence was found suggesting that recommender systems are a risk factor in the committing of intimate image abuse offences. However, where intimate images have been posted by a perpetrator on a U2U services that has a public newsfeed, there is a risk that content recommender systems may process, rank and disseminate those images to other users' feeds, amplifying the number of users who see these images. The likelihood of posted intimate images being disseminated may increase if they receive sufficient

---

[1011] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[1012] Thorn (Wolak, J. and Finkelhor, D.), 2016. Sextortion: Findings from a survey of 1,631 victims. [accessed 4 August 2023].
[1013] Deepfakes are a specific type of media that involves the use of AI algorithms, particularly generative AI models, to create and/or modify videos, images or audio to create realistic synthetic content. This can be done by superimposing the face of a person onto the body of another person in a video or image as well as voice manipulation with lip syncing. Generative AI models can also create entirely new synthetic content, including sexual synthetic content of existing individuals. Deepfakes are shared as user generated content on user-to-user services but could also potentially be created using functionalities present on user-to-user services. The technology used for this purpose has many legitimate and beneficial applications, but we are concerned here only with its misuse in the generation of deepfake intimate images.
[1014] Sensity AI, 2020. Automating Image Abuse: Deepfake bots on Telegram. [accessed 15 February 2024].
[1015] Glitch, 2023. AI Deepfake Roundtable 1. [accessed 17 January 2024].
[1016] Glitch, 2023.
[1017] Rahman-Jones, I., Taylor Swift deepfakes spark calls in Congress for new legislation, BBC News, 27 January [accessed 19 November 2024].

engagement in the form of likes, shares, and comments. Once in an open channel of communication, there is also the risk of such images being disseminated via direct messaging features.

Risk factors: Business models and commercial profile

6.90 There is no known evidence specific to business models and the facilitation or commission of these offences.

7. Extreme pornography offence

Warning: this chapter contains content that may be upsetting or distressing, including examples of sexually violent acts.

Summary analysis for the extreme pornography content offence: how harm manifests online, and risk factors

Material considered 'extreme pornography' can include assault, rape and violence. There is limited evidence on the possession of extreme pornographic content, for several reasons, including the ethical and legal limitations on conducting research into it. However, some evidence suggests that there is a link between extreme pornographic material and CSAM, with a common pathway to CSAM online being the consumption of legal, and then increasingly problematic and potentially extreme pornography. We therefore draw similarities in risks where appropriate.

Service type risk factors:

User-to-user pornography services that provide user-generated pornography may be at a higher risk of showing or setting out that they offer extreme pornographic content.

User base risk factors:

Insights about demographic risks tend to concern perpetrators rather than victims or survivors. Crime data indicates that the creation or posting of extreme pornographic content, which suggests also the *viewing* of this content, is primarily committed by men, which indicates that **gender** – in relation to having a male-skewed user base – could be a risk factor.

Functionalities and recommender systems risk factors:

Extreme pornography can be facilitated by **posting content**, in this case, images and videos, on user-to-user (U2U) services. The ability to **search for user-generated content (UGC)** on U2U services may also help users find extreme pornography content.

Additionally, other functionalities can be involved in the perpetration of this offence. **Hyperlinks** may also take a user from a U2U service with legal content to services with more extreme and potentially illegal content; this could include extreme pornographic content.

Inferences from similar offences such as CSAM (see the Child sexual exploitation and abuse chapter, indicate that the ability to **download content** enables users to store and view extreme pornographic content and to share it with others. **Anonymous profiles** also give perpetrators confidence that they can avoid detection and are likely to increase the risk of extreme pornographic content being present online.

Content recommender systems also appear to play a role in suggesting increasingly extreme pornographic content to users. **User groups** can facilitate users viewing and sharing extreme pornography, with users exchanging content with like-minded individuals. It is possible that the ability to **edit visual media** can lead to the creation of realistic-looking deepfake extreme pornographic content.

Livestreaming could allow users to broadcast extreme pornographic content in real-time. Being able to **post goods and services for sale** can amplify the risk posed by live streaming, as livestream sessions can be selected and purchased by users.

Business model risk factors:

The revenue models of some online user-to-user pornography services rely on ensuring a supply of new content to maintain and increase their user base.

This applies to **advertising and subscription-based revenue models**, which can incentivise these services to allow content to be uploaded in the most 'friction-free' manner to maximise user engagement and minimise the cost of moderation. These services may, consequently, be less able to effectively detect and moderate extreme pornographic content.

Introduction

7.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the extreme pornography offence listed under 'Relevant offences'; and
- the use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

7.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm, which we discuss as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual or encountered in combination with content of a different kind. Where appropriate, we also consider the wider societal effects of the harm caused to individuals because of exposure to the content or activities that amount to relevant offences. In this case, this would relate to how extreme pornographic content might normalise and/or promote harmful or violent sexual behaviour.

Relevant offences

7.3 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Concerning extreme pornography, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

7.4 In this chapter, we consider the following offence:

- possession of extreme pornography[1018]

7.5 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of this offence.

7.6 'Extreme pornography' is an umbrella term used in UK law to cover several categories of images which are illegal to possess. Although the legislation varies slightly across legal systems in the UK, extreme pornography broadly covers images which are produced principally for sexual arousal, and which depict extreme or obscene behaviours. Possession involves having 'custody or control' over the content.

7.7 Extreme pornographic content includes realistic and explicit depictions of necrophilia, bestiality, acts threatening a person's life, acts that could result in serious injury to specific parts of the body, rape and assault by penetration.

7.8 For more details on the offence and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

How the extreme pornography offence manifests online

7.9 This section is an overview which looks at how the extreme pornography offence manifests online, and how individuals may be at risk of harm.

7.10 To put the risks of harm into context, a study into the policing of extreme pornography analysed 591 cases across England and Wales regarding the charging and recording of this offence between 2015 and 2017.[1019] [1020] The most charged category was that of extreme pornography involving an animal.[1021] This is likely to be because it is easier to identify these illegal images as extreme pornographic material.[1022]

7.11 Moreover, in 2018 to 2019 there were 28 prosecutions for possession of an extreme pornographic image portraying rape or assault by penetration.[1023]

7.12 However, the true scale of extreme pornographic content available online is challenging to quantify. Evidence indicates it is far more widespread than prosecution numbers would indicate. Research published by the UK government in 2020 noted there is little proactive regulation of extreme pornography, which means that internet users may be able to access unlawful material on legal pornography sites.[1024]

[1018] In England, Wales and Northern Ireland, this falls under section 63 of the Criminal Justice and Immigration Act 2008.
[1019] McGlynn, C. and Bows., H., 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, *The Journal of Criminal Law*, 83(6). [accessed 18 November 2024].
[1020] The study used data from 591 cases obtained from 33 police forces across England and Wales regarding charging and recording of this offence between 1st April 2015 and 31st March 2017. 254 recorded incidents in 2015-16 and 337 for 2016-17.
[1021] McGlynn, C. and Bows., H. 2019.
[1022] McGlynn, C. and Bows., H. 2019.
[1023] Crown Prosecution Service, 2019. Violence against women and girls report 2018-19. [accessed 4 September 2023].
[1024] UK Government, 2020. The relationship between pornography use and harmful sexual behaviours. [accessed 05 November 2024].

7.13 The report also notes that that sexually explicit content that was previously 'harder to access' – for example content depicting violence – is now much more easily available.[1025] This is echoed in research from Durham University that found that 12%, or one in every eight titles, of analysable content on the landing pages of the top three user-to-user pornography services in the UK described sexual activity that constituted sexual violence.[1026] Some of the content in both of these reports may overlap with extreme pornographic content.

7.14 As this is an image-based offence, any service that allows the uploading and sharing of images or videos could, in principle, be used to commit or facilitate the offence.

7.15 Due to gaps in the literature, Ofcom has assessed how similar content-driven harms manifest online to identify risk factors that could play a role in extreme pornography. As with other illegal content, we presume that extreme pornography could manifest through posting images or videos onto U2U services. Once posted to these services it can be viewed – either intentionally or unintentionally – by other users. Some users may actively and knowingly seek out this content while others may be unaware they are accessing illegal material.[1027] Users can also download this content to their own devices or share it on other services. This illegal content can also be broadcast in real time, such as through livestreaming.

7.16 As with child sexual abuse material (CSAM) (see the Child sexual abuse and exploitation (CSEA) chapter for more information), hyperlinks or plain-text URLs to extreme pornographic content are also a way in which users can share extreme pornography.

Risks of harm to individuals presented by the extreme pornography offence

7.17 Research on the risks of harm from extreme pornographic content can be difficult to measure, including legal and ethical challenges faced by researchers aiming to study material that is illegal.[1028] Given these limitations, this section provides an overview on the risks of harm presented by this content drawing on both existing research on extreme pornography, as well as looking more broadly at evidence exploring link between harm and possession and circulation of sexually explicit material.

7.18 In the development of the extreme pornography offence, aspects of the legislation on these categories of images have been underpinned by the need to address the harm done to those depicted in the content, and to address the wider societal harm caused by circulation of these images. Research on this topic indicates that extreme pornographic material can

[1025] Upton, J., Hazell, A., Abbott, R. & Pilling, K. (The Behavioural Architects on behalf of the Government Equalities Office and Women and Equalities Unit), 2020. The relationship between pornography use and harmful sexual behaviours. [accessed 29 October 2024].

[1026] For their study the authors used the World Health Organisation definition of sexual violence, which is broader than the legal threshold for extreme pornography. They focused on four broad categories of sexual violence: sexual activity between family members; aggression and assault; image-based sexual abuse and coercive and exploitative sexual activity. Source: Vera-Gray, F., McGlynn, C., Kureshi, I., Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology,* 61(5), pp.1-18. [accessed 18 November 2024].

[1027] Vera-Gray, F., McGlynn, C., Kureshi, I., Butterby, K., 2021.

[1028] Jones, S. and Mowlabocus, S., 2009. Hard Times and Rough Rides: The Legal and Ethical Impossibilities of Researching 'Shock' Pornographies, *Sexualities,* 12(5), pp. 613-628. [accessed 18 November 2024].

normalise the acts depicted.[1029] This includes depictions of gender-based violence, including sexual violence, rape and sexual assault, as well as high risk sexual behaviours such as explicit and realistic depictions of life-threatening injury.

7.19 Similar research looking at depictions of sexual violence argues that the availability of this content can contribute to harmful and coercive 'sexual scripts' about what behaviours and attitudes are acceptable or pleasurable.[1030] This can have a range of impacts connected to normalisation, including that users may be less likely to understand what sexual acts are unlawful or harmful. It can contribute to sexual violence being more easily dismissed or going unreported.[1031] [1032]

7.20 In addition, there is some evidence directly linking consumption of violent pornography (including that which may not meet the definition of extreme pornography) to violent behaviour, however, this has been heavily scrutinised.[1033] Several studies have identified weaknesses in the research base that link exposure to extreme pornography with violent behaviour. Meta-analysis studies in other fields have found minimal effects of exposure to violent content leading to violent behaviour, for example in gaming.[1034] Research from the UK Government notes that those perpetrating harmful sexual behaviours are influenced by multiple factors which could include viewing pornography, but it is never one factor alone that leads to this behaviour.[1035]

7.21 There is also some evidence to suggest that there may be a link between possessing and viewing extreme pornography and viewing child sexual abuse material. There have been cases of people being charged with offences related to indecent images of children alongside the extreme pornography offence.[1036] [1037] This suggests that there may be a coalescence of problematic online behaviour; an individual who engages with extreme pornography online might also actively seek out other illegal content online. This could

---

[1029] McGlynn, C. and Bows., H., 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, *The Journal of Criminal Law*, 83(6). [accessed 18 November 2024].

[1030] Ethan A. Marshall, Holly A. Miller, Jeffrey A. Bouffard, PhD. 2018. Bridging the Theoretical Gap: Using Sexual Script Theory to Explain the Relationship Between Pornography Use and Sexual Coercion, Journal of Interpersonal Violence, 36(9-10). [accessed 29 October 2024].

[1031] During the development of legislation on extreme pornography, the Home Office stated that it is *"possible that such material may encourage or reinforce interest in violent and aberrant sexual activity to the detriment of society as a whole".* Source: Home Office, 2005. Consultation: On the possession of extreme pornographic material. [accessed 4 September 2023].

[1032] Researchers have argued that the availability of extreme pornographic content, including rape and non-consensual sexual penetration, sustains a culture in which sexual violence is not only not taken seriously, but risks creating a culture in which it is normalised. Source: Vera-Gray, F., McGlynn, C., Kureshi, I., Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology,* 61(5), pp.1-18. [accessed 29 October 2024].

[1033] McGlynn, C. and Rackley, E., 2009. Criminalising extreme pornography: a lost opportunity, *Criminal law review*, 4, pp. 245-260. [accessed 29 October 2024].

[1034] Drummond, A., Sauer, J. D. and Ferguson, C. J., 2020. Do longitudinal studies support long-term relationships between aggressive game play and youth aggressive behaviour? A meta-analytic examination. *Royal Society Open Science,* 7:200373. [accessed 29 October 2024].

[1035] Upton, J., Hazell, A., Abbott, R. & Pilling, K. (The Behavioural Architects on behalf of the Government Equalities Office and Women and Equalities Unit), 2020. The relationship between pornography use and harmful sexual behaviours. [accessed 29 October 2024]

[1036] Antoniou, A. and Akrivos, D., 2017. *The Rise of Extreme Porn—Legal and Criminological Perspectives on Extreme Pornography in England & Wales* cited in McGlynn, C. and Bows., H., 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, *The Journal of Criminal Law*, 83(6). [accessed 29 October 2024].

[1037] Gilbody-Dickerson, C., 2023. Adam Britton: What we know about man who pleaded guilty to 'grotesque' sexual abuse of dozens of dogs, *i¸26 September.* [accessed 27 September 2023]

suggest a link between the two offences, but the sample is too small to draw any definitive conclusions.

7.22 The impact on those involved in producing extreme pornographic content, consensually or non-consensually, is outside the scope of the extreme pornography offence.

7.23 However, the production of the content is tied to the consumption of the content online. As mentioned in the Register of Risks chapter 'Child sexual abuse and exploitation (CSEA)', if people watch this material online, the market will supply it, which can exacerbate the harm to those involved in its production. This is similarly applicable to the production of extreme pornographic content, such as content depicting serious injury, threats to life, rape and sexual assault. This is related to the risk of serious harm arising from high-risk behaviours such as explicit and realistic depictions of suffocation and strangulation.[1038] It also includes any depictions of acts at or above the threshold of assault – consent cannot be given in these cases.

7.24 In addition, extreme pornographic content depicting rape or sexual assault is evidently harmful to the victim or survivor. Impacts include severe harm to that person's physical and mental health, including trauma, depression and anxiety. The knowledge of such content existing may also have long-lasting effects on an individual including through re-traumatisation.

7.25 There are also scenarios of individuals who have created content simulating sexual assault and rape consensually but who could also be harmed in its making. For instance, some actors in the pornography sector have expressed negative physical and mental effects because of producing pornographic content.[1039] [1040]

Evidence of risk factors on user-to-user services

7.26 We consider that the risk factors below are likely to increase the risks of harm relating to the extreme pornography offence. This is also summarised at the start of the chapter.

Risk factors: Service types

7.27 As an image-based offence, any service that allows the uploading and sharing of images or videos could, in principle, be used to commit or facilitate the offence. Research indicates that user-to-user pornography services can be used to commit or facilitate extreme pornography offences.

User-to-user pornography services

7.28 There is evidence to show that user-to-user pornography services which provide user-generated pornography are a risk factor. A study of sexual violence in mainstream online pornography found that 12% of first-time viewers of the landing pages of mainstream adults services in the UK described a sexual activity that constituted sexual violence, as

[1038] McGlynn, C. and Woods, L. 2022. Pornography and Online Safety Bill. [accessed 29 October 2024]
[1039] Cole, S., 2020. A new wave of reckoning is sweeping the porn industry. *VICE*, 10 June. [accessed 4 September 2023].
[1040] In some of these cases the violent act being filmed may not meet the legal threshold for extreme pornographic content. But it is possible to assume that if actors experience boundary violation and non-consensual acts during the filming of more mainstream pornographic content, it is likely that actors involved in the making of extreme pornographic content are also likely to be affected.

defined by the study.[1041] The British Board of Film Classification (BBFC),[1042] found that in a study of young people and pornography, most respondents said they were exposed to upsetting or disturbing videos (usually related to violent or aggressive pornography) for the first time through "*videos appearing on homepages of pornography sites or as a suggested video".*[1043]

7.29 The study of the landing pages of three online UK user-to-user pornography services found that content describing criminal acts was regularly presented there. This included content describing criminal offences including rape. The authors of the study stress that these landing pages are what first-time users see when accessing these services, and act as a 'shop window' for young people and others who are new to the world of online porn.[1044] This study was based on the descriptions of the content, rather than analysing the content itself. While it is unclear whether the content itself represented a realistic depiction of these offences, the study highlights the prevalence of descriptions of extreme pornography. Although this study focused on user-to-user pornography services, it is possible that these findings could be relevant to other U2U services that allow pornographic content on their services.

Messaging Services

7.30 Messaging services have been used by perpetrators to share extreme pornographic content. There have been cases of perpetrators using messaging sites to share self-generated extreme pornographic content.[1045] There have also been instances of users sending unsolicited extreme pornographic content through messaging apps.[1046]

Risk factors: User base

User base demographics

7.31 The following section outlines primary evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

7.32 Data suggest that gender may be a risk factor in the commission or facilitation of an extreme pornography offence. Analysis of recorded incidents of extreme pornography by 33 police forces found that the vast majority of those charged were men (97%). It is

---

[1041] Vera-Gray, F., McGlynn, C., Kureshi, I., Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology,* 61(5), pp.1-18. See How the extreme pornography offence manifests online section for more information.

[1042] The BBFC is the UK's regulator of film and video. It is responsible for the 18 classifications for sex work, and the R18 category for legally-restricted content, primarily for explicit works of consenting sex or strong fetish material involving adults.

[1043] BBFC, 2020. Young people, pornography and age-verification. [accessed 6 September 2023].

[1044] Vera-Gray, F., and McGlynn, C., 2021. Sexually violent pornography is being promoted to first-time users of top sites. [accessed 6 September 2023].

[1045] News.com.au, 2023. 'I can't stop. I don't want to': Dog rapist sent disturbing Telegram messages about sordid urges, 26 September. [accessed 26 September].

[1046] Zaccaro, M. and PA Media, 2023. Met Police officer jailed over extreme pornographic image, BBC News, 17 March. [accessed 27 September 2023]. Sharman, D., 2023. Police probe 'digital sex abuse' after female journalists sent extreme porn, *HoldtheFrontPage.co.uk*, 27 September. [accessed 27 September 2023].

therefore reasonable to assume that more men than women may commit offences relating to extreme pornography.[1047]

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles

7.33 Anonymity can facilitate the commission of the extreme pornography offence. If a service allows for the creation of anonymous user profiles, this removes friction and may lead to a user feeling more confident in posting illegal content such as extreme pornographic content.

User networking

User groups

7.34 We have not identified evidence specifically linking extreme pornography to the ability to create user groups. However, as previously mentioned, in a context where extreme pornography also appeals to users seeking CSAM material, it may be possible that user groups are used in similar ways to CSAM (see Child sexual abuse and exploitation (CSEA)). User groups would therefore facilitate the extreme pornography offence as they would allow content to be shared with users who have similar interests.

User communications

Direct messaging and encrypted messaging

7.35 Direct messages are a channel through which extreme pornographic content can be shared. For example, perpetrators have shared extreme pornographic content on encrypted messaging sites. [1048] Perpetrators may choose to use encrypted messaging to avoid detection.

Livestreaming

7.36 We have not found specific evidence of livestreaming being used in the commission or facilitation of the extreme pornography offence but consider that it would be possible for livestreaming to be used. As identified in the Register of Risks chapters 'Child sexual abuse and exploitation (CSEA)' and 'Terrorism', livestreaming provides a space for users to stream potentially harmful material in a seamless and sometimes unmonitored way. It would be possible for livestreaming to be used to broadcast sexual acts that, if captured or recorded, may potentially constitute possession of extreme pornography.

Posting content (image, video)

7.37 The ability to post content, in this case, images and videos, is an important functionality in the commission or facilitation of the extreme pornography offence. In 2020, Pornhub, an online user-to-user pornography service, removed 10 million videos, amounting to about 80% of its content, after high-profile coverage raised concerns about the availability of illegal material, including CSAM and non-consensually shared intimate images, hosted on

[1047] McGlynn, C. and Bows., H. 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, *The Journal of Criminal Law*, 83(6). [accessed 29 October 2024].
[1048] News.com.au, 2023. 'I can't stop. I don't want to': Dog rapist sent disturbing Telegram messages about sordid urges, 26 September. [accessed 26 September].

the service.[1049] Christian Action, Research and Education (CARE)[1050] said that this action showed that *"large U2U services are unsure and are unable to record the levels of illegal and extreme material hosted on their services*."[1051] There is a risk that services which attract adult content, particularly where illegal content has already been found, may also be the types of services where extreme pornography is found.[1052]

Transactions and offers

Post goods or services for sale

7.38 Functionalities that allow users to post goods or services for sale could be used to commit or facilitate extreme pornography offences, as it is possible that users could use these functionalities to advertise access to extreme pornographic content. This tactic is used in the facilitation of CSAM offences, and we think it reasonable to assume that it could also facilitate the extreme pornography offence (see Child sexual abuse and exploitation (CSEA) chapter).

Content exploring

User-generated content searching

7.39 A U2U service's search function could allow a user to find extreme pornographic content. Vera-Gray and McGlynn's study of sexually violent content on UK user-to-user pornography services found that pornography which contravened the service's terms and conditions could be found through simple keyword searches.[1053] This suggests that the search functions on online user-to-user pornography services could allow words associated with acts that could be considered extreme pornographic content. Although users searching for the terms may not necessarily be looking for illegal content, such functions would facilitate a user in discovering extreme pornographic content.

Building lists or directories

7.40 As with in-service search functionalities, many online user-to-user pornography services currently use lists and directories to help users identify content. It is reasonable that these functionalities could also be used by those seeking to identify extreme pornographic content on services, by searching categories within directories that are linked to extreme pornographic terms. The curation and categorisation of images will likely increase the discoverability of such material.

Hyperlinking

7.41 Hyperlinks that take users to other services could direct users to channels which offer less mainstream pornographic material than the content hosted on the initial service. Such links may take a user from a U2U service with legal content to services with more extreme and potentially illegal content. These hyperlinks can facilitate a user's exposure to extreme

[1049] Concerns were not necessarily raised about extreme pornography material, but primarily child sexual exploitation and abuse material and intimate image abuse (IIA).
[1050] CARE is a Christian public policy charity based in the United Kingdom.
[1051] CARE response to 2022 Ofcom Call for Evidence: First phase of online safety regulation.
[1052] This is not a specific observation about potentially illegal and extreme content currently available on Pornhub. However, Pornhub's actions help us draw an inference that extreme pornographic content may exist on U2U services.
[1053] Vera-Gray, F. and McGlynn, C., 2021. Sexual Violence in Mainstream Online Pornography. [accessed 6 September 2023].

pornography and provide a relatively easy user journey to sites showing more niche or even extreme pornographic content.

7.42 However, the use of hyperlinks to access illegal content is not limited to online user-to-user pornography services. As evidenced in the chapter 'Child sexual abuse and exploitation (CSEA)' (see CSAM), hyperlinks and plain-text URLs linking to illegal images are shared among perpetrators on a variety of service types, allowing potential perpetrators to access and download extreme pornographic content.

Content storage and capture

Downloading content

7.43 There is currently limited evidence in the public domain on the extent to which users are downloading this content. However, it is reasonable to assume that this functionality could be used by people to download extreme pornographic content.

Content editing

Editing visual media

7.44 The creation of deepfake material[1054] is a rapidly emerging technology that is likely to have implications in the creation and possession of extreme pornographic material. Although there is currently a lack of direct evidence, it is reasonable to assume that such technology could be used to create realistic-looking, extreme pornographic content which could then be shared on U2U services. This could be achieved by creating an entirely new image or editing other images. The Illegal Content Judgement Guidance (ICJG) has more information on when an altered image should be considered illegal content.

Recommender systems

Content recommender systems

7.45 Evidence suggests that the design of recommender systems could be a risk factor in the facilitation and commission of extreme pornographic content offence. There is a growing body of research that indicates that services hosting user-generated pornographic content seek to maximise user retention as needed by their business model. This is explored more further in this chapter. McGlynn and Woods suggest that this can result in limited content moderation processes which can allow the presence of extreme content available for dissemination by recommender systems, where used. They note that "*the suggestions of 'related content' - aiming at user retention – may push increasingly extreme content*." This indicates that recommender systems can play a role in suggesting increasingly extreme pornographic content.[1055]

7.46 A BBFC study of how young people discover and engage with pornography online found that several boys had found themselves getting 'lost' in an online world of pornography

[1054] Deepfakes are a specific type of media that involves the use of AI algorithms, particularly generative AI models, to modify videos, images or audio to create realistic synthetic content. This is often done by superimposing the face of a person onto the body of another person in a video or image as well as voice manipulation with lip syncing. Deepfakes are shared as user generated content on user-to-user services but could also potentially be created using functionalities present on user-to-user services. Deepfake technology is currently used to create content that can be harmful; however, we acknowledge that it may also have positive use cases.

[1055] McGlynn, C., and Woods, L., 2022. Image-Based Sexual Abuse, Pornography Platforms and the Digital Services Act. [access 6 September 2023].

where they found increasingly violent videos. [1056] This content may not necessarily constitute extreme pornographic content, but the evidence indicates how the design of recommender systems can lead to users being exposed to increasingly extreme content. We think it is reasonable to assume that where extreme pornographic content is present on a service and has not been detected and taken down by content moderation processes, this content may be recommended to users who are not necessarily looking for it directly.

7.47 Moreover, as detailed in Risk Factors: Service Types, landing pages of online user-to-user pornography services have been found to show content describing criminal acts, and recommender systems play a role in determining or influencing what is shown on a landing page.

Risk factors: Business models and commercial profiles

Revenue models

Advertisement-based revenue models

7.48 Some services that use the dissemination of user-generated pornographic content have an incentive to enable the posting of videos or images in the most 'friction-free' way, with low levels of moderation and due to their advertising revenue model which relies on increasing their user base and user retention. This makes such services less incentivised to detect and moderate extreme pornographic content, suggesting that the advertising business model can be a risk factor. A report by the Centre to End All Sexual Exploitation states that some services strive to make the uploading of content a friction-free experience to maintain a content offering that will continue to attract users, and that this could result in instances of extreme pornography being presented.[1057]

7.49 We note that payment process providers may exert financial pressure on services to take down extreme pornographic content. Evidence suggests that credit card companies can ban customers from using its service to purchase on a pornographic website with such content.[1058]

[1056] BBFC, 2020. Young people, pornography and age-verification. [accessed 21 September 2023].
[1057] Centre to End All Sexual Exploitation, 2021. Expose Big Porn. [accessed 21 September 2023].
[1058] Mohan, M., 2020, Call for credit card freeze on porn sites, *BBC News*, 8 May. [accessed 21 September 2023]; Goodwin, J., 2020. Visa continues its ban on Pornhub but will allow payments on some of its parent company's sites, *CNN*, December 23. [accessed 21 September 2023].

8. Sexual exploitation of adults

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for sexual exploitation of adults offences: how harm manifests online, and risk factors

This chapter analyses the risks of harm arising from offences relating to the sexual exploitation of adults – this includes adults who have been forced into sex work as well as consenting adult sex workers who are being exploited. In this chapter, we use the terms 'adult sex worker' or 'victim and survivor' rather than 'prostitute' and 'prostitution', unless referencing the legislation that uses those terms. This is to align with widely accepted terminology conventions which seek to better reflect the experiences and dynamics in this area.

The International Labour Organization estimates that 6.3 million people globally are currently experiencing forced commercial sexual exploitation and that nearly four out of five are women or girls. We recognise that the same characteristics identified as risk factors in sexual exploitation of adults can at times also be safety measures for adult sex workers. For example, while social media services, marketplaces and listings services have been identified as potential risk factors there is also evidence that advertising sexual services online is generally safer than soliciting in public places.

The risk of harms to individuals from sexual exploitation of adults offences include both physical and psychological effects. Victims and survivors may suffer from threats of physical abuse, rape or sexual violence. This can also have an effect on their mental health, with effects including anxiety, depression, self-harm and post-traumatic stress disorder. The risk factors identified below may lead to individuals experiencing the risks of harm from these offences.

Service type risk factors:

Social media services are likely to be used by potential perpetrators to recruit victims and to advertise the services of the victims and survivors they have gained control of for sexual exploitation. **Marketplaces and listings services** are also used to advertise services.

Messaging services, particularly those with encryption, can also be used by potential perpetrators to communicate with victims and survivors.

User base risk factors:

Gender and **age of users** are risk factors - there is evidence that women and younger people are more vulnerable to exploitation. Other user base demographics can also be risk factors; individuals with **intellectual disabilities**, **language barriers** or who are **homeless** are more vulnerable to exploitation.

Functionalities and recommender systems risk factors:

The ability to **post goods or services for sale**, such as through advertisements, also enables perpetrators to advertise and broadcast the sexual services of adults in exploitative environments. **Encrypted messaging** is also a functionality that can be used by buyers and abusers as a tool to arrange the transaction of services, while **user profiles** can also be used to identify individuals. These functionalities enable the commission of the offence 'controlling a prostitute for gain'.

Other functionalities also present risk of harm from this offence. **Livestreaming** can be used by perpetrators to advertise and broadcast exploitation, with these streams reaching a large global base of potential consumers. Evidence suggests that the ability to **post or send location information** can also be used to identify and target individuals. **Direct messaging** is used for communication between perpetrators, buyers, and victims and survivors.

Business model risk factors

Services that **generate revenue through advertising** can be at risk, as offenders may be able to use adverts to lure victims, who can then be coerced or controlled into sexual activities, and it has been recognised that some online services have profited from selling adverts for sexual services provided by potential victims and survivors of this kind of coercion and control.

Introduction

8.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the sexual exploitation offences listed under 'Relevant offences'; and
- the use of these services for the commission and/or facilitation of these offences (collectively, the 'risks of harm').

8.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

8.3 Sexual exploitation is the inducement of a commercial sex act generally by means of force, fraud or coercion. We have addressed various aspects of sexual offences over several chapters in an attempt to reflect the risk and victim impact in a focused and proportionate way. This chapter covers the sexual exploitation of adults. Trafficking offences, including sexual exploitation of both adults and children, are discussed in the chapter Human trafficking. Offences relating to child sexual exploitation and abuse ('CSEA') are discussed in the Register of Risks chapters 'Child sexual exploitation and abuse (CSEA)', 'Grooming' and 'Child sexual abuse material (CSAM)' chapters.

Relevant offences

8.4 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Concerning the sexual exploitation of adults, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

8.5 This chapter will consider the available evidence on the following offences which aim to target those who recruit others into prostitution for their own, or someone else's, gain:

- Causing or inciting prostitution for gain[1059]
- Controlling a prostitute for gain[1060]

8.6 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and in relation to offences in Scotland, being involved art and part in the commission of these offences).

8.7 Everyone's experience is unique, and the offences will affect individuals differently. This chapter covers two distinct experiences. For the first offence (a), this chapter will consider the online experiences of individuals who are coerced or forced into sex work; these individuals could be victims of trafficking and/or controlled by another person. For the second offence (b), this chapter will consider the online experience of adults who identify as consenting sex workers who may experience harms linked to being controlled or exploited by another person.

8.8 'Sexual exploitation of adults' will be used in this chapter as an umbrella term for these offences. The evidence provided in this chapter may capture a broader range of issues and harm than the specific elements of the offences contained within the Act, but they have been included as we consider that they will help services in understanding the risk of harm from offences relating to the sexual exploitation of adults. For example, we have used evidence that refers to sexual exploitation among both adults and children as some of the research in this area which contributes to our understanding does not focus solely on adult sexual exploitation. In some cases, it considers both issues together and in others does not make a distinction between adult sexual exploitation and CSEA.

8.9 For more details on these offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

How sexual exploitation of adults manifests online

8.10 This section is an overview which looks at how sexual exploitation of adults offences manifest online, and how individuals may be at risk of harm.

8.11 To put the risks of harm into context, the internet has enabled those seeking to exploit sex workers or to coerce people into sex work. The International Labour Organization estimates

[1059] Section 52 of the Sexual Offences Act 2003; Article 62 of the Sexual Offences (Northern Ireland) Order 2008 (S.I. 2008/1769 (N.I. 2)).
[1060] Section 53 of the Sexual Offences Act 2003; Article 63 of the Sexual Offences (Northern Ireland) Order 2008 (S.I. 2008/1769 (N.I. 2)).

that 6.3 million people are currently experiencing forced commercial sexual exploitation and that nearly four out of five are women or girls.[1061]

8.12 However, the working practices of consenting adult sex workers have been fundamentally altered by the internet. UK adult sex workers have reported that the development of online sex work has improved their safety, enabling them to screen potential clients, and allowing them to work more independently.[1062]

Risks of harm to individuals presented by offences relating to the sexual exploitation of adults

Causing or inciting prostitution for gain

8.13 The facilitation of this offence online generally involves an offender trying to exploit individuals with opportunities that end up involving them in being sexually exploited. For example, an advertisement offering accommodation in exchange for sex.

8.14 There is evidence of people who have been forced or tricked into sex work by other individuals, sometimes traffickers. These individuals can use fraudulent job opportunities or initiate romances to recruit and incite children or adults into forced sex work. These tactics are often traded against the promise of shelter, material possessions, transport or drugs, for example.[1063]

8.15 Individuals who are seeking to exploit others can also use pre-existing relationships to take advantage of the victim, or the victim's relationship with someone else. Common pre-existing relationships include connections on online services such as social media services, spouses or intimate partners, mutual friends, friends or classmates, drug dealers, parents or legal guardians, religious leaders, extended family including partners of a parent or guardian, landlords, employers or teachers.[1064]

8.16 Coercion features heavily in the manifestation of these offences. Physical restraint may form a part of this abuse, but in many cases, coercion manifests largely in other ways, including but not limited to withholding pay, physical abuse, threats of physical abuse, and rape or sexual violence.

8.17 Evidence has found that those who have been sexually exploited have been subjected to physical, sexual and psychological violence. This can lead to harm to victims' and survivors' physical health, including HIV infections, gynaecological problems, substance and alcohol abuse and long-term physical injury.[1065] It also has an impact on their mental health, with effects including anxiety, depression, self-harm and post-traumatic stress disorder.[1066] Sexual exploitation can also create negative issues for relationships and caregiving.

[1061] International Labour Organization (ILO), 2022. Global Estimates of Modern Slavery: Forced Labour and Forced Marriage. [accessed 5 July 2023].
[1062] Beyond the Gaze (Sanders, T., Scoular, J., Pitcher, J., Campbell, R. and Cunningham, S), 2018. Beyond the Gaze: Summary Briefing on Internet Sex Work. [accessed 5 July 2023].
[1063] Human Trafficking Institute, 2021. Federal Human Trafficking Report 2020. [accessed 5 July 2023].
[1064] Human Trafficking Institute, 2021.
[1065] McQuaid, J., 2020. Understanding the psychological effects of sex trafficking to inform service delivery, *Forced Migration Review*. [accessed 5 July 2023].
[1066] McQuaid, J., 2020.

Controlling a prostitute for gain

8.18 The manifestation of this offence online might be the advertising of sexual services from an individual, posted by someone who is controlling that individual and forcing them to provide the sexual services. The facilitation of such an offence might be an individual messaging a buyer and arranging the sale of sexual services from a person who is being forced to provide those services.[1067]

8.19 There is evidence to suggest that the transition of sex workers to using online advertising can reduce the risks to them. The National Police Chiefs' Council stated that selling sexual services online is generally safer than soliciting in public places, for example.[1068] However, harms can, and do, still, arise.

8.20 Consenting sex workers can also be the victims of sexual exploitation. Some, or all, aspects of their work may be controlled by a third person or persons for gain. This could manifest as another individual controlling which clients the sex worker has to engage with, or controlling the money the sex worker earns. Adult sex workers who are victims of this sexual offence can be forced to work long hours for minimal pay and be threatened with violence if they do not adhere.[1069]

8.21 As with victims and survivors who are forced into sex work, adult sex workers who are being controlled can face similar risks of violence and damage to their physical and mental health.

Evidence of risk factors on U2U services

8.22 We consider that the risk factors below are likely to increase the risks of harm relating to the sexual exploitation of adults. This is also summarised at the start of the chapter.

Risk factors: Service types

8.23 Research indicates that the following types of services can be used to facilitate or commit offences related to the sexual exploitation of adults: marketplaces and listings services,[1070] social media services, video-sharing services, and messaging services.

Social media services, marketplaces and listings services

8.24 In terms of identifying and recruiting individuals to sexually exploit, social media services and online marketplaces are identified as some of the most common types of services used by traffickers in a report from the United Nations Office on Drugs and Crime.[1071] This is likely to also be true for victims and survivors of the sexual offences addressed in this chapter. These services provide potential perpetrators with an opportunity to post false or misleading opportunities to recruit their targets and usually provide the means for initiating communication.[1072]

[1067] National Police Chiefs Council, 2024. Sex Work National Police Guidance. [accessed 14 October 2024]
[1068] National Police Chiefs Council, 2024.
[1069] Stop the Traffik, 2021. Sex Work and Exploitation: What do you need to know?. [accessed 5 July 2023].
[1070] While our evidence only names online marketplaces, we expect similar risks of harm to arise from listings services due to similarities in the characteristics typically found on these service types.
[1071] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons 2020. [accessed 5 July 2023].
[1072] United Nations Office on Drugs and Crime, 2020.

8.25 Social media services provide perpetrators with a large pool of potential victims and the ability to collect personal information from that individual and connect with them quickly. The role of service types in the identification and recruitment of victims and survivors for exploitation is explored further in the Human trafficking chapter.

8.26 In terms of controlling an individual who is being sexually exploited, marketplaces and listings services provide an opportunity for perpetrators to place advertisements for the sexual services of someone they are sexually exploiting. A United Nations report found "*the analysis of court cases report that regular online marketplace sites, on which anyone can post or browse advertisements to sell or buy any service (from job vacancies to the sale of equipment, cars and clothes), are being used to advertise services obtained from victims of human trafficking.*"[1073] This will probably also be the case for other victims and survivors who are being sexually exploited.

Messaging services

8.27 Messaging services can be used in the sexual exploitation of adults. A study from Thorn, an organisation researching technology and sexual abuse, showed that respondents who were being sexually exploited, in this case by traffickers, reported the trafficker communicating with buyers using certain messaging services.[1074] There is additional research to suggest that messaging services with encryption are a risk factor.[1075]

Video-sharing services

8.28 Our evidence shows that livestreaming, a functionality that is common to video-sharing services, is a risk factor. Livestreaming has been found to be used for acts of exploitation in the commission of the offence of 'controlling a prostitute for gain'.[1076]

Risk factors: User base

User base demographics

8.29 The following section outlines key evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

8.30 Data suggests that user base characteristics including **age, gender, disability, language, media literacy** and **socio-economic background** of users could lead to an increased risk of harm to individuals.

8.31 The age of users can be a risk factor in the sexual exploitation of adults. One study highlighted the factors associated with adolescent commercial exploitation as sexual victimisation, younger age drug and alcohol abuse, being a victim of intimate partner abuse and a sense of sexual stigmatisation. This study also noted a link between the cessation of

[1073] United Nations Office on Drugs and Crime, 2020.
[1074] Thorn, 2018. Survivor Insights: The role of technology in domestic minor sex trafficking. [accessed 5 July 2023].
[1075] See Risk factors: functionalities and recommender systems section for more information.
[1076] See Risk factors: functionalities and recommender systems section for more information.

commercial sexual exploitation and the completion of higher education, suggesting that lower academic attainment may also be a risk factor.[1077]

8.32 The gender of users is also a risk factor in the sexual exploitation of adults; women are disproportionately likely to be sexually exploited. The International Labour Organization estimates that 6.3 million people are currently experiencing forced commercial sexual exploitation, and that nearly four out of five are women or girls.[1078] The National Police Chiefs' Council reported that 91% (more than 9 in 10) of sexual exploitation impacts women and girls.[1079]

8.33 Evidence suggests that language barriers could make people more vulnerable to people seeking to sexually exploit them.[1080]

8.34 Media literacy skills could also be a factor, as users with lower levels of critical understanding could find it more difficult to identify false or misleading opportunities that are exploitative.

8.35 Individuals from disadvantaged socio-economic backgrounds could be more at risk of being sexually exploited. Poverty and homelessness are possible reasons why people may find themselves vulnerable to being sexually exploited.[1081] Similarly, an individual's immigration status can make them more vulnerable to sexual exploitation. The impact of unstable immigration status was identified as a significant contributor to increasing vulnerability to exploitation in a recent large-scale evidence review into the risks of modern slavery in the UK among children and young adults. This is, in part, due to a reluctance to engage with relevant support services such as health or law enforcement due to a fear of what might happen due to their immigration status.[1082] [1083]

8.36 Disability can be a risk factor in the commission of sexual exploitation of adult offences. Individuals who are neuro-divergent, or have illness, disability or health conditions are more vulnerable to exploiters. They can be viewed as 'easy targets' as their care and support needs may affect their ability to protect and defend themselves.[1084]

[1077] Reid, J. 2014. Risk and resiliency factors influencing onset and adolescence-limited commercial sexual exploitation of disadvantaged girls, *Criminal Behaviour and Mental Health*, 24 (5), p.332-344.
[1078] ILO, 2022. Global Estimates of Modern Slavery: Forced Labour and Forced Marriage. [accessed 5 July 2023].
[1079] National Police Chiefs' Council, 2023. Violence Against Women and Girls: Strategic Threat Risk Assessment. [accessed 5 July 2023].
[1080] Preventing Exploitation Toolkit, n.d. Communication difficulties. [accessed 5 July 2023].
[1081] Preventing Exploitation Toolkit, n.d. Communication difficulties. [accessed 5 July 2023].
[1082] The Rights Lab/ECPAT UK (Celiksoy, E., Schwarz, K., Sawyer, L., Gorena, P. V., Ciucci, S., Yin, S. & Durán, L.) 2024. Prevention and identification of children and young adults experiencing, or at risk of, modern slavery in the UK. [accessed 28 October 2024].
[1083] International Organization for Migration (IOM), 2019. Handbook on Protection and assistance to migrants vulnerable to violence, exploitation and abuse. [accessed 13 November 2024].
[1084] Preventing Exploitation Toolkit, n.d. Communication difficulties. [accessed 5 July 2023].

Risk factors: functionalities and recommender systems

User identification

User profiles and fake user profiles

8.37 User profiles are used by those seeking to sexually exploit others as a recruitment tool. Perpetrators will use the information users provide in their user profiles to identify vulnerable people and create user profiles in order to engage with potential victims online. They can be used to facilitate the sexual exploitation of adults, in particular, the offence of 'causing or inciting prostitution for gain', as user profiles allow exploiters to target individuals who could be vulnerable with fraudulent promises of opportunity. This is explained in more detail in the Unlawful immigration and Human trafficking chapters.

8.38 It is possible that potential perpetrators use fake user profiles to falsely present themselves to the person they are exploiting or intending to exploit. A fake user profile could add legitimacy to an individual who is misrepresenting themselves, for example through the presence of a profile picture or personal information. This legitimacy can also be used by an individual to portray themselves as offering legitimate services that are in fact sexual exploitation.

User communication

Livestreaming

8.39 Livestreaming is used for acts of exploitation in the commission of the offence of 'controlling a prostitute for gain.' A United Nations report found that the internet is being used to exploit individuals (adults and children), using the broadcasting or livestreaming of acts of exploitation. Such streams reach large bases of potential consumers across the world, as has been evidenced in cases.[1085]

Direct messaging and encrypted messaging

8.40 Direct messaging can be used to facilitate the offence of 'controlling a prostitute for gain' as it allows buyers to arrange the purchase of exploitative services. Buyers communicate with exploiters using messaging functionalities on social media services and private messaging services.[1086] This use of direct messaging to arrange purchases is a growing trend.

8.41 The added security and privacy offered by encrypted messaging enables abusers, having contacted a victim, to entice, manipulate, entrap and exploit them into sexual activities for their profit. Using encrypted messaging means that there is less chance for moderation and detection of this activity. A UN report found that the use of multiple services by traffickers shows that they are aware of the risk of monitoring or surveillance, so they often move their communication from open groups on social media services to encrypted or anonymised services such as a private messaging service.[1087] This suggests that those seeking to sexually exploit others may use encrypted messaging in their commission of the offence of 'controlling a prostitute for gain.'

8.42 Although direct messaging functionality is a risk factor in the sexual exploitation of adults, this same functionality is an important safety measure for adult sex workers. Exchanging

[1085] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons 2020. [accessed 5 July 2023].
[1086] Thorn, 2018. Survivor Insights: The role of technology in domestic minor sex trafficking. [accessed 5 July 2023].
[1087] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons 2020. [accessed 5 July 2023].

messages with a client allows an adult sex worker to assess the safety concerns they have about the client before meeting in person.

Posting or sending location information

8.43 Functionalities that allow users to post or send location information are used as tools to identify vulnerable people. These functionalities can also be used to monitor people's movement and can be used to control others. This is explained in more detail in the chapter 'Controlling or coercive behaviour'.

8.44 The ability to post or send location information can also be used as a safety measure by consenting adult sex workers, as it can make other people aware of their real-time location, should this be needed.[1088]

Transactions and offers

Posting goods or services for sale

8.45 The ability to post goods or services for sale plays a primary role in the commission or facilitation of the offence of 'controlling a prostitute for gain'. While the evidence that follows can refer to 'advertising,' it is likely to include content which has the effect of marketing or promoting goods and services, rather than just paid advertising, which is expanded on in the revenue model section below.

8.46 Posting goods and services for sale as a means of advertising them is used as a way for exploiters to set up a 'shop window' for clients to choose who they would like to 'buy.' These listings are likely being placed on open channels, to maximise the number of people who see the post. In one court case quoted in a United Nations report, a perpetrator connected one victim-survivor with more than 100 sex buyers over 60 days using 'online advertisement'.[1089] It is unclear whether this was a consequence of paid advertising (see Revenue model section below), or a posting that advertised the victim-survivor.

8.47 There is a wide body of literature on the indicators (warning signs) of sexual exploitation of adult sex workers. A common analytical tool used by many police forces in the United Kingdom is the Sexual Trafficking Indicator Matrix (STIM), which was created by academics in 2021 to support law enforcement officers in identifying adult services website profiles potentially used by human trafficking offenders.[1090] For example, one study using the STIM tool was by Giommoni and Ikwu, who analysed over 17,000 advertisements for female sex workers that were *"posted on the largest dedicated platform for sex work services in the UK."* [1091] This, along with the author's description of these advertisements,[1092] suggests that

---

[1088] Beyond the Gaze, 2018. Safety and Privacy for Online Sex Workers. [accessed 5 July 2023].
[1089] United Nations Office on Drugs and Crime, 2020.
[1090] L'Hoiry, X., Moretti, A. & Antonopoulos, G. A. 2021. Identifying sex trafficking in Adult Services Websites: an exploratory study with a British police force, *Trends in Organised Crime*. Vol. 27. [accessed 13 November 2024].
[1091] These include: the use of third- or first-person plural pronouns; the same phone number used in more than one advertisement; a high degree of similarity between sex workers' advertisements; sex workers offering risky or violent sexual services; advertisements promoting inexpensive sex services; sex workers moving frequently between several locations; sex workers moving to a different location along with other sex workers; sex workers offering in-call services only; advertisements using words alluding to the youthful characteristics of the sex workers; and stating a dress size typical of underage women.
[1092] Most advertisements in the study are said to provide 'demographic information about the sex worker (e.g. town where they are active, nationality, age, etc.), information on their physical appearance (e.g. height, hair and eye colour, etc.), sexual orientation (e.g. bi-sexual, heterosexual, etc.), sexual services provided, and pricing. Moreover, most

this includes posted content (as opposed to paid advertising) which, in turn, has the effect of advertising a service by offering it for sale.

8.48 The study established a set of ten indicators of human trafficking and found that most of the advertisements (58.3%) contained one indicator, 21.3% (more than 1 in 5) of the advertisements presented two indicators and 1.7% of advertisements (nearly 2 in 100) reported three or more indicators of human trafficking.[1093] Although this study focused on victims and survivors who had been sexually exploited via trafficking, it is likely these same indicators will also be relevant in identifying people who are being sexually exploited but who have not necessarily been trafficked.

8.49 In summary, these studies show that advertising services by posting them for sale is an important way in which individuals can 'control a prostitute for gain'. However, the advertising of sex services by a third party can act as a safety measure for some consenting adult sex workers. This may apply to adult sex workers who have difficulty advertising their services themselves; for example, those whose first language is not English or those who do not have access to technology or technical literacy. In such cases, a third party can advertise and risk-assess the clients on the sex worker's behalf.

8.50 This functionality can also facilitate the commission of the offence of 'causing or inciting prostitution,' as posts which advertise services are used as a tool by exploiters to recruit victims. CPS recognises that there has been an increase in reports of "*advertisements posted on classified advertising websites*" where landlords offer accommodation in exchange for sex. The CPS notes that "*such arrangements can lead to the exploitation of highly vulnerable persons who are struggling to obtain accommodation.*"[1094] [1095]

8.51 Online censorship and restrictions may worsen the experiences of sex workers and may lead to different forms of exploitation. A report by Hacking//Hustling looked at the consequences of the FOSTA-SESTA legislation in the United States in 2018. The stated goal of this law was to reduce human trafficking, and it was argued the law put increased pressure on Internet platforms to censor their users. Consequently, the sex worker communities that this law directly impacted claimed it has increased their exposure to violence and left those who rely on sex work as their primary form of income without many of the tools they had used to keep themselves safe.[1096] [1097]

---

advertisements have free text spaces that the sex workers use to introduce themselves, a public and private gallery to post pictures and videos, and an 'interview' section where sex workers provide more details about themselves and their services.'

[1093] Giommoni, L. and Ikwu, R., 2021. Identifying human trafficking indicators in the UK online sex market, *Trends in Organized Crime*. [accessed 19 September 2024].

[1094] Crown Prosecution Office, 2019. Prostitution and Exploitation of Prostitution. [accessed 20 September 2023].

[1095] See also the Register of Risks chapter 'Human trafficking' for information on related offences.

[1096] Blunt, D. & Wolfe, A. 2020. Erased: The Impact of FOSTA-SESTA. [accessed 11 September 2024].

[1097] See also Albert, K., et al. 2021. FOSTA in Legal Context. [accessed 26 November 2024]. Chamberlain, L. 2019. FOSTA: A Hostile Law with a Human Cost. [accessed 26 November 2024].

Risk factors: Business models and commercial profiles

Revenue models

Advertising-based models

8.52 Services that generate revenue through advertising may enable perpetrators to reach potential victims and customers of those engaged in the sexual exploitation of adults. The opportunity to advertise can be used by perpetrators, including sexual traffickers, to entice victims into a situation where they are captured, controlled and coerced into sexual activities.

8.53 This is supported by a United Nations report[1098] setting out that services which provide classified listings or other advertising provide an effective mechanism for traffickers to place attractive but false career opportunities to entrap victims into being trafficked.[1099]

8.54 In addition, a study, 'The Role of Technology in Domestic Minor Sex Trafficking' recognised that some services which offer online advertising have been *"profiting from prostitution ads*".[1100] We think that a significant proportion of this is likely to have been presented by trafficked sexual workers.

[1098] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons 2020. [accessed 5 July 2023].
[1099] *"Examples of advertisements used to attract victims often include wording that describes the possibility of living a luxurious life or promising jobs in industries such as modelling or entertainment".* Source: United Nations Office Drugs and Crime, 2020.
[1100] Thorn, 2018. Survivor Insights: The role of technology in domestic minor sex trafficking, p38 [accessed 5 July 2023].

9. Human trafficking

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for human trafficking offence: how harm manifests online, and risk factors

Human trafficking encompasses a wide range of harmful activities. It can involve modern slavery, and victims and survivors include adults and children. It is estimated that there were 122,000 people living in modern slavery in the UK in 2021.

Notable forms of human trafficking where harm can manifest online include sexual exploitation and abuse, forced labour, and criminal exploitation such as county lines exportation of illegal drugs.

An individual's experience of harms from human trafficking offence are unique to their situation, but victims and survivors of these offences can be considered one of the most vulnerable groups at risk of complex mental health difficulties, as well as long-lasting physical health problems.

Service type risk factors:

Social media services and **messaging services** are shown to be risk factors for human trafficking offence. Among other elements, they are used to target potential victims and advertise services. Perpetrators can identify their victim on social media before transitioning to private messaging services to further the facilitation of the offence.

Marketplace and listing services, particularly adult services websites are recognised as risky in the context of the human trafficking, sexual exploitation and abuse in particular. This is because these types of websites are often used to advertise services provided by victims who have been trafficked and are being coerced.

Functionalities and recommender systems risk factors:

User profiles, including fake profiles used to hide the perpetrator's real identity and manipulate a victim/survivor, can be exploited by a perpetrator looking to build trust with their victim. **Posting content** is also used to lure victims and promote illegal services in human trafficking offence.

Direct and encrypted messaging are a risk factor as they are used by traffickers to recruit and communicate with victims, as well as for communication between those involved in exploitation.

Business model risk factors:

Services providing classified listings or other advertising opportunities may increase risk, as the services are incentivised to maximise advertising revenues. The opportunity to advertise can be used by traffickers to reach and attract potential victims, who can then be lured into a situation where they are captured, controlled and coerced. Where sex workers are victims of human trafficking, online listings that advertise sex workers' services may also present a risk of facilitating human trafficking.

Introduction

9.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to the human trafficking offence listed under 'Relevant offences'; and
- the use of these services for the commission and/or facilitation of this offence (collectively, the 'risks of harm').

9.2 We set out the characteristics of U2U services that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

9.3 The evidence presented below focuses on the human trafficking offence on U2U services. It uses the terms 'trafficking' and 'modern slavery' to describe relevant offences.[1101] It is important to recognise that these are harms which are also experienced by children. Other, related offences such as unlawful immigration and sexual exploitation are discussed in other chapters.[1102]

9.4 Ofcom has reviewed the available evidence to develop this assessment of the risks of the human trafficking offence and how they manifest on U2U services. It may not be an exhaustive account of the possible uses of online services to facilitate this offence.

Relevant offences

9.5 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to human trafficking, Ofcom is required to consider the risks of harm connected with the priority offence listed in Schedule 7 of the Act:

[1101] 'People smuggling' and 'people trafficking' are different concepts in law. Offences relating to 'people smuggling' will generally relate to the Immigration Act offences, whereas 'people trafficking' will generally be offences under the Modern Slavery Act, Crown Prosecution Service, Updated 6 July 2022. Source: Crown Prosecution Service, 2022. Modern Slavery, Human Trafficking and Smuggling. [accessed 25 September 2023].

[1102] See the Register of Risks chapters 'Unlawful immigration' and 'Child sexual exploitation and abuse (CSEA)'.

a) arranging or facilitating the travel of another person, or taking a relevant action, with a view to them being exploited (human trafficking)[1103]

9.6 The Act also covers inchoate offences such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and, in relation to offences in Scotland, being involved art and part in the commission of those offences).

9.7 The human trafficking offence covered in this chapter involves the exploitation of children or adults in some way. Specifically, it is an offence for a person to take a "relevant action" with a view to another person being exploited. "Relevant actions" are:

- the recruitment of another person,
- the transportation or transfer of another person,
- the harbouring or receiving of another person,
- the exchange or transfer of control over another person, or
- the arrangement or facilitation of any of these actions.

9.8 The human trafficking offence may involve, or take place alongside, a wide range of abuses and other criminal offences such as grievous bodily harm, assault, rape or child sexual abuse.

9.9 We recommended readers also review the Register of Risks chapters on Child sexual exploitation and abuse (Grooming and CSAM) and the Sexual exploitation of adults for further details on sexual exploitation.

9.10 Online aspects of the human trafficking offence can entail potential victims being proactively targeted by perpetrators, as well as advertisements for job opportunities to recruit people into being trafficked. People are trafficked for numerous reasons, including sexual, labour and criminal exploitation. This chapter will include evidence of the role of online services in human trafficking, regardless of whether the person is trafficked for slavery, labour, criminal or sexual exploitation.[1104]

9.11 For more details on the offence and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

How human trafficking offence manifest online

9.12 This section is an overview which looks at how human trafficking offence manifest online, and how individuals may be at risk of harm.

9.13 Human trafficking offence take a range of forms. Examples include forced labour, domestic servitude,[1105] forced criminal activity, sexual abuse and exploitation, as well as organ

[1103] Section 2 of the Modern Slavery Act 2015; section 1 of the Human Trafficking and Exploitation (Scotland) Act 2015 (asp 12); section 2 of the Human Trafficking and Exploitation (Criminal Justice and Support for Victims) Act (Northern Ireland) 2015 (c. 2 (N.I.)).

[1104] Some of the evidence in this chapter draws on evidence from outside of the UK context, where this is relevant to build understanding of the risks of harm people face, as well as the role of online services in facilitating offences.

[1105] Lovett, S., 2023. Foreign diplomats trap and abuse domestic workers in private households across London. The Telegraph, 2 February. [accessed 14th August 2024].

harvesting.[1106] The concept of modern slavery is used to describe many kinds of human trafficking, indicating the harmful as well as often-hidden nature of these offences.

9.14 As victims and survivors are often concealed and/or unable to escape their circumstances, it is difficult to measure the prevalence of human trafficking or modern slavery.[1107] Global estimates suggest that in 2021 there were 28 million people around the world in forced labour and 22 million in forced marriage.[1108] Save the Children have said that *"children account for 27% of all human trafficking victims worldwide, and two in three child victims are girls"*.[1109] Children are more easily placed in situations of danger by smugglers and forced into participation in criminal activities.[1110]

9.15 In 2023 there were 17,004 potential victims of modern slavery in the UK referred to the Home Office, including 7,432 children. There were also 4,929 adults who were suspected to be adult victims but would not consent to referral. All of these were the highest annual figures of their kind since the National Referral Mechanism (NRM) began in 2009.[1111]

9.16 It is, however, likely that the prevalence of human trafficking is much higher than this. The 2023 Global Slavery Index estimates there were 122,000 people living in modern slavery in the UK in 2021.[1112]

9.17 There are many ways in which trafficking offences are committed, commissioned and facilitated online, but a few common and sometimes overlapping patterns stand out. One is for a trafficker to groom a victim by establishing a relationship with them before exploiting them via manipulation, blackmail, or the threat of and/or use of violence. Another is to share information about a supposed work or financial opportunity to attract a victim, before engaging in exploitation. Finally, some trafficking victims have their services advertised online – for example, in the case of trafficked sex workers – something which can potentially expand the scope of traffickers' activities if they are able to earn more money.

9.18 The relationship between traffickers and victims or survivors varies widely. Perpetrators may have pre-existing relationships with potential victims/survivors via friends and family,[1113] purposefully create a personal relationship to commit a trafficking offence (for example, by presenting themselves as a potential friend or romantic partner)1114 or trick

---

[1106] Reported cases of organ harvesting in the UK are rarer, but examples include 3 people found guilty of exploiting a vulnerable victim for illegal organ harvesting in 2023. Source: Crown Prosecution Service, 2023. Updated with sentence: Senior Nigerian politician jailed over illegal UK organ-harvesting plot. [accessed 23 September 2024]

[1107] ONS, 2020. Modern slavery in the UK: March 2020. [accessed 12th August 2024].

[1108] International Labour Organisation, 2022. Global Estimates of Modern Slavery: Forced Labour and Forced Marriage. [accessed 13 August 2024].

[1109] Save the Children, n.d. The Fight Against Child Trafficking. [accessed 25 September 2023].

[1110] UNODC, 2019. Children on the move, smuggling and trafficking. [accessed 25 September 2023].

[1111] Potential victims of modern slavery in the UK who come to the attention of authorised first responder organisations are referred to the National Referral Mechanism (NRM). Authorised first responder organisations include local authorities, specified non-governmental organisations (NGOs), police forces and specified government agencies. Adults (aged 18 or above) must consent to being referred to the NRM, whilst children under the age of 18 need not consent to being referred. Home Office, 2024. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2023. [accessed 12th August 2024].

[1112] Walk Free, 2023. Global Slavery Index: United Kingdom. [accessed 17 October 2024].

[1113] For example, government guidance about county lines notes that in cases of criminal exploitation, victims sometimes have pre-existing connections with perpetrators. Source: Home Office, 2023. Criminal exploitation of children, young people and vulnerable adults: county lines. [accessed 17 October 2024].

[1114] Council of Europe, 2022. Online and technology-facilitated trafficking in human beings. [accessed 17 October 2024].

someone looking for an employment opportunity into thinking they are providing non-exploitative work for them.[1115]

9.19 The role of U2U services varies depending on the context, but they are used for both targeting victims and for communicating with them during exploitation, or with others involved in exploitation. For example, in one case from 2022, four people were convicted of human trafficking offence after recruiting victims aged 14 to 17 on social media to carry out theft and fraud on their behalf.[1116]

9.20 For more detail on how other offences closely related to human trafficking manifest online, readers should refer to the Register chapters 'Sexual exploitation of adults', and 'Child sexual abuse and exploitation (CSEA)'.

Risks of harm to individuals presented by the human trafficking offence

9.21 Human trafficking offence present multiple risks to both children and adults; the victims and survivors can be entrapped, trafficked or exploited in many different ways. [1117] [1118] Depending on the severity of their exploitation, perpetrators might be violent towards victims or take control of their finances.[1119] In some cases, victims and survivors are deprived of necessities like food or coerced into exploitation via threats to family members.[1120]

9.22 Whilst victims and survivors come from a range of backgrounds, the gendered nature of some aspects of human trafficking mean it is an area where there is often violence against women and girls.[1121] In addition, there can be a relationship between socio-economic factors and human trafficking, with victims and survivors more vulnerable to exploitation depending on their financial needs.

9.23 Victims and survivors of human trafficking are considered to be one of the most vulnerable groups at risk of "complex mental health difficulties, including anxiety, depression, aggression, suicidal ideation and post-traumatic stress disorder (PTSD)".[1122]

[1115] UNODC, 2020. Global Report on Trafficking in Persons 2020: Chapter 5. [accessed 17 October 2024].
[1116] Brown, E., 2022. Dozens of teenage girls trafficked in UK modern slavery first, Unilad, 10 February. [accessed 17 October 2024].
[1117] Children are trafficked for many of the same reasons as adults including, but not limited to, forced marriage, domestic servitude, forced labour, organ harvesting, criminal exploitation and sexual exploitation and can experience one of or a multitude of forms of abuse and exploitation. Source: NSPCC, n.d. Child Trafficking. [accessed 25 September 2023].
[1118] The CPS discusses exploitation and talks about extensive examples in its guidance here. Source: CPS, 2022. Modern slavery, human trafficking and smuggling. [accessed 25 September 2023].
[1119] National Crime Agency, n.d. Modern slavery and human trafficking. [accessed 17 October 2024].
[1120] Unseen UK, n.d. Frank's story. [accessed 17 October 2024].
[1121] For example, the majority of referrals under the National Referral Mechanism for potential sexual exploitation are for women and girls. Source: Home Office, 2024. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, quarter 2 2024 (April to June). [accessed 17 October 2024]. In addition, there is evidence of links between experiences of violence among sex workers and having experiences of being trafficked. Source: Deering, K.N., Amin, A., Shoveller, J., Nesbitt, A., Garcia-Moreno, C., Duff, P., Argento, E. and Shannon, K., 2014. A systematic review of the correlates of violence against sex workers. American Journal of Public Health, 104(5). [accessed 17 October 2024].
[1122] Unseen (Garbers, K., Malpass, A., Saunders, L., Horwood, J., McLeod, H., Anderson, E. and Farr., M.), 2021. Impact of mobile technology for survivors of modern slavery and human trafficking. [accessed 25 September 2023]; Hestia, 2023. Underground lives: Forgotten children - The intergenerational impact of modern slavery. [accessed 17 October 2024].

9.24 Trafficking can have both short- and long-term effects on children, including damage to their physical and mental health, limited or no access to education, and drug and alcohol use.[1123]

9.25 The risks of harm faced by victims and survivors of human trafficking vary depending on how they are trafficked and for what purpose. We explore some of the more common types of trafficking: sexual exploitation, criminal exploitation, labour exploitation, and trafficking involving migrant victims.[1124]

Sexual exploitation and abuse

9.26 There is evidence of people who have been forced or tricked into sex work by other individuals, including in pornography available online, as well as in escort or prostitution services advertised online.[1125] These types of exploitation are often gendered; at a global level, forced commercial sexual exploitation predominantly targets women and girls.[1126]

9.27 Traffickers can use fraudulent job opportunities, target vulnerabilities or initiate romances to recruit children or adults into sex trafficking. These tactics are often traded against the promise of shelter, material possessions, transport or drugs.[1127] Individuals who are seeking to exploit others can also use pre-existing relationships to take advantage of the victim/survivor or the victim's/survivor's relationship with another person. Common pre-existing relationships include social media contacts, spouses or intimate partners, mutual friends, friends or classmates, drug dealers, parents or legal guardians, religious leaders, extended family, including partners of a parent or guardian, landlords, employers, or teachers.[1128] There are also examples of individuals who have formed a relationship with someone seeking to traffic them for purposes of sexual exploitation.[1129]

9.28 Evidence has found that those who are sexually exploited may be subject to physical, sexual and psychological violence.[1130] This can lead to harm to their physical health, including HIV infections, gynaecological problems, substance and alcohol abuse and long-term physical injury.[1131] It also has an effect on their mental health; effects include anxiety, depression, self-harm and post-traumatic stress disorder (PTSD).[1132] Sexual exploitation can also create issues for relationships and caregiving. More information on risks associated with sexual exploitation can be found in chapter Sexual exploitation of adults offences.

[1123] NSPCC, 2023. Protecting children from trafficking and modern slavery. [accessed 25 September 2023].
[1124] It is worth noting that many victims may experience multiple types of trafficking. This is captured in National Referral Mechanism data about potential modern slavery victims.
[1125] All-Party Parliamentary Group on Commercial Sexual Exploitation, 2023. Pornography Regulation: the case for Parliamentary Reform. [accessed 17 October 2024]. All-Party Parliamentary Group on Commercial Sexual Exploitation, 2021. Bust the Business Model: How to stop sex trafficking and sexual exploitation in the UK. [accessed 17 October 2024]. Home Affairs Committee, 2023. Oral Evidence: Human Trafficking. [accessed 17 October 2024].
[1126] International Labour Organization, 2022. Global Estimates of Modern Slavery: Forced Labour and Forced Marriage. [accessed 17 October 2024].
[1127] Human Trafficking Institute, 2021. Federal Human Trafficking Report 2020. [accessed 25 September 2023].
[1128] Human Trafficking Institute, 2021.
[1129] In one case a human trafficking survivor met their exploiter via a dating app. Source: Ashcraft, E., 2019. Human trafficking survivor shares how she fell victim after Tinder date, KSL News, 23 January. [accessed 17 October 2024].
[1130] An example of a human trafficking survivor who was subjected to sexual abuse and violence is discussed in this news report. Source: Deas, A. and Mortimer, H., 2022. I was trafficked, raped, and left for my abusers to find, BBC News, 27 September. [accessed 17 October 2024].
[1131] McQuaid, J. 2020. Understanding the psychological effects of sex trafficking to inform service delivery. [accessed 25 September 2023].
[1132] McQuaid, J. 2020.

9.29 Child sexual exploitation is a form of child sexual abuse. It occurs when an individual or group takes advantage of an imbalance of power to coerce, manipulate or deceive a child or young person under the age of 18 into sexual activity. This is done: (a) in exchange for something the victim needs or wants, and/or (b) for the financial advantage or increased status of the perpetrator or facilitator.[1133]

9.30 Sexual exploitation of a child is complex, with no one linear manifestation either online or offline. In the context of human trafficking, child sexual exploitation and abuse can be its express purpose, for example in cases of child sex trafficking,[1134] or something that takes place alongside other kinds of trafficking, such as in some cases of child criminal exploitation.[1135]

9.31 A study into child sex trafficking in North America highlights a range of evidence sources where children having been recruited online via services including social media, listings and gaming services. Traffickers' strategies often involved "*initiating interpersonal relationships or even deceptively posing as an old friend*." Children became entrapped and enmeshed in exploitation via tactics including violence, threats, and traffickers encouraging dependency to them based on drug addiction, acting as a sole source of money or basic needs, and victims becoming pregnant.[1136]

9.32 Potential victims of modern slavery in the UK in 2023 included 1,119 children who were referred due to concerns about sexual exploitation.[1137] In addition to those, a further 330 cases identified sexual exploitation as part of a wider group of harms experienced by the child victims. This included forced labour, domestic servitude, and organ harvesting.[1138] Child sexual exploitation does not happen exclusively within sex trafficking; further manifestations are explored below in the Criminal exploitation section and in chapter Child sexual exploitation and abuse (CSEA) offences.

Criminal exploitation including county lines

9.33 There is a significant and growing body of evidence of criminals using online services to target individuals who they wish to exploit by persuading or coercing them into participating in or supporting criminal activities.

9.34 Forced criminal activity can take numerous forms, including acquisitive crimes such as shoplifting, and forced labour in illegal activities such as the production and selling of drugs. Perpetrators may force, coerce or groom individuals into money laundering or possession

---

[1133] Department for Education, 2017. Child sexual exploitation. [accessed 25 September 2023].
[1134] Thorn, 2019. Survivor Insights: The Role of Technology in Domestic Minor Sex Trafficking. [accessed 17 October 2024]; Thorn, 2024. Survivor Survey. [accessed 17 October 2024]; Reuters, 2022. Former UK army officer jailed for online child sex abuse. [accessed 17 October 2024].; Baird, K., Connolly, J., 2021. Recruitment and Entrapment Pathways of Minors into Sex Trafficking in Canada and the United States: A Systematic Review. *Trauma, Violence & Abuse*, 24(1). [accessed 28 October 2024].
[1135] See for example, Pearson, J., Cavener, J. 2024. Professionals' understanding of the County Lines phenomenon: Insights from a study exploring the perceptions of young peoples' supported accommodation staff. *Children and Youth Services Review*, vol 156. [accessed 28 October 2024].
[1136] Baird, K., Connolly, J., 2021. Recruitment and Entrapment Pathways of Minors into Sex Trafficking in Canada and the United States: A Systematic Review. *Trauma, Violence & Abuse*, 24(1). [accessed 28 October 2024]
[1137] Home Office, 2024. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2023. [accessed 12 August 2024]. These figures include 447 potential victims where more than one type of exploitation was highlighted in the referral.
[1138] Independent Anti-Slavery Commissioner, 2021. Child trafficking in the UK 2021: a snapshot. [accessed 25 September 2023].

of illegal items. Victims can be coerced or groomed into involvement, or not be aware they are involved in criminal activity.[1139]

9.35 An example of criminal exploitation from an NCA investigation in 2023 found three suspected victims/survivors of human trafficking working at a cannabis farm in Stroud, Gloucestershire. The investigation officer described how they had "*uncovered a criminal network believed to be involved in setting up cannabis farms and staffing them with the victims of modern slavery and human trafficking*".[1140]

9.36 The term 'county lines' refers to an illicit enterprise involving urban street gangs and organised criminal networks in the UK that export illegal drugs to one or more areas within the country.[1141] Children and vulnerable adults can be exploited into supporting these activities, via manipulation and violence.[1142] In 2023, 1,559 county lines referrals were flagged as part of the NRM, accounting for 9% (nearly 1 in 10) of all referrals received in the year. Most of these referrals were for boys aged under 17.[1143]

9.37 Children are sometimes targeted specifically for forced participation in criminal activities, such as drug dealing through 'County Lines' and financial fraud. Evidence suggests online services have been used in a numerous ways to facilitate in child criminal exploitation.[1144] One study has suggested that this has taken place in a context where 'newer' gangs have used social media "*to brand themselves and to appear attractive for recruits and customers alike*."[1145] Research with children and parents into experiences of child criminal exploitation in Wales points to a number of factors that can increase vulnerability to exploitation, including the promise of financial gain, peer influence, and the experience of transitions between different educational settings.[1146] A report from Barnardo's also talks about school holidays being a period when children may be more vulnerable to criminal exploitation.[1147] One victim/survivor of criminal exploitation has talked about being targeted online at a time when they wanted to run away from home, and that perpetrators took advantage of him based on his being less likely to be stopped by police.[1148]

9.38 The Children's Society have noted that based on their work delivering support to victims of child criminal exploitation, children have been targeted in a range of locations in the UK and may be coerced into distributing drugs in their local area by someone based in another part of the country altogether – indicating the potential for acts of exploitation to take place entirely online. In addition, they note instances where child sexual abuse material (CSAM)

[1139] Hestia (Papadaki, H.), 2020. Underground Lives: Criminal Exploitation of Adult Victims. [accessed 18 October 2024].
[1140] NCA, 2023. NCA targets crime group suspected of operating slave labour cannabis farms. 25 January. [accessed 25 September 2023].
[1141] National Crime Agency, n.d. County Lines. [Accessed 12 May 2023].
[1142] Maxwell, N. and Wallace, C., 2021. Child Criminal Exploitation in Wales. [accessed 17 October 2024].
[1143] Home Office, 2024. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2023. [accessed 12th August 2024].
[1144] Children and Young People's Centre for Justice, 2023. Understanding Child Criminal Exploitation in Scotland: A Scoping Review. [accessed 1st November 2024].
[1145] Whittaker et. Al, 2020. No two gangs are alike: The digital divide in street gangs' differential adaptations to social media. [accessed 1st November 2024].
[1146] Maxwell, N. and Wallace, C., 2021.
[1147] Barnardo's, 2023. Invisible Children: Understanding the risk of the cost-of-living crisis and school holidays on child sexual and criminal exploitation. [accessed 17 October 2024].
[1148] Campbell, C., 2021. County lines gang 'recruited teen in 80 minutes via Snapchat', BBC News, 14 April. [accessed 17 October 2024].

of victims is made, and where online 'collateral' (for example, incriminating messages) can be used by perpetrators to ensure compliance.[1149]

9.39 The wider effects of this exploitation are physical and sexual violence, emotional abuse resulting in a decline in emotional wellbeing, and exposure to adverse childhood experiences, leading to the likelihood of engaging in risky behaviours such as alcohol abuse, underage and/or unprotected sex, teenage pregnancy, illicit drug consumption and becoming a victim of, or committing, a violent act.[1150]

9.40 Other wider effects, commonly associated with such exploitation, stem from the risk of the child becoming indebted to an exploiter. Debt bondage is commonplace, and can lead to further exploitation in county lines, threats to the child and their family, serious injury, sexual exploitation to pay off the debt with sexual favours, pressure to commit other crimes such as robbery, burglary and fraud, and having siblings or friends being drawn into county lines to help pay off the debt.[1151] Children exploited in this way have been seen to exhibit traits such as self-harm, a significant decline in school attendance, a decline in emotional wellbeing and a reduction in pro-social connections.[1152]

9.41 Across different forms of criminal exploitation, there can be severe effects on mental health, with survivors describing feelings of shame or guilt, as well as challenges maintaining relationships with family and friends due to the stigma associated with criminal activity.[1153] There are also physical risks being involved in criminal or gang activity; a report into child criminal exploitation notes the deaths and serious harm experienced by children drawn into criminal exploitation, from bladed weapons and firearms.[1154]

Forced labour and labour exploitation

9.42 In 2023, potential labour exploitation victims accounted for 34% (more than one-third) of all referrals of adults into the NRM.[1155] Some of these cases include labour exploitation where victims/survivors are involved in explicitly criminal activities as described in 'organised crime', but others can relate to more conventional industries where the business/employer may or may not be knowingly involved in criminal offences.[1156] Examples cited by the NCA include car washes, nail bars, construction, sea fishing, and agriculture.[1157] Other sectors where people have been involved in potential cases of labour exploitation include recruitment agencies[1158] and hospitality.[1159]

[1149] Children's Society response to the Protection of Children May 2024 Consultation.
[1150] Public Health England, 2021. County Lines exploitation: applying All Our Health. [accessed 12 May 2023].
[1151] Rescue and response, 2019. Rescue and response county lines project. [accessed 12 May 2023].
[1152] Public Health England, 2021.
[1153] Hestia (Papadaki, H.), 2020. Underground Lives: Criminal Exploitation of Adult Victims. [accessed 18 October 2024].
[1154] Children and Young People's Centre for Justice, 2023. Understanding Child Criminal Exploitation in Scotland: A Scoping Review. [accessed 1st November 2024].
[1155] Home Office, 2024. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2023. [accessed 12 August 2024].
[1156] A relevant example in this news report involved a gang who forced 16 victims to work at a branch of McDonald's and a factory. The gang took nearly all of the victims' pay and housed them in poor living conditions. Source: McLennan, W., Shepka, P. and Ironmonger, J., 2024. McDonald's and supermarkets failed to spot slavery, BBC News, 30 September. [accessed 18 October 2024].
[1157] UK Parliament, 2023. Written evidence submitted by the NCA. [accessed 17 October 2024].
[1158] GLAA, 2023. Nepalese recruitment agency directors handed slavery order. [accessed 17 October 2024].
[1159] Europol, 2024. 51 persons arrested in crackdown on labour exploitation. [accessed 17 October 2024].

9.43 The Metropolitan Police have explained how victims of labour exploitation can be forced to work very long hours for little or no pay, and are often kept, and work, in terrible conditions.[1160] The charity Unseen's helpline encountered cases of potential modern slavery in the care sector, where workers recruited from abroad had their possessions such as phones and passports confiscated by their employers, given poor living conditions, and in one case made to work for longer than 24 consecutive hours without breaks. They also describe hearing about issues with potential debt bondage, where workers are charged likely illegal fees to work in the UK, and then have deductions made from their wages to repay the debt.[1161]

Trafficking offences and migrant victims

9.44 Whilst victims of trafficking can be of any nationality or citizenship status, there are some trafficking risks that are more unique to migrants.[1162] Some people have been trafficked to another country directly for purposes such as sexual and labour exploitation, with online services being used to target victims by providing false advertisements of opportunities to work abroad.[1163] [1164] The role of organised crime groups in both trafficking and smuggling means that some irregular migrants are put at risk of trafficking when attempting travel to and/or upon arrival in their destination country.[1165] Europol note that *"irregular migrants can end up exploited for the debt of the smuggling fees, either by the criminal networks facilitating their travels or by other criminal groups with which the smugglers cooperate."*[1166]

9.45 This is borne out in statistics about irregular migration, which show that 1 in 10 people who have arrived in the UK on a small boat have been referred as potential victims of modern slavery.[1167]

9.46 Evidence also points to trafficking risks being heightened for some people without formal citizenship status in the UK. Exclusion from mainstream employment and welfare systems means vulnerability to labour exploitation can be heightened.[1168] There are also cases of non-UK nationals with leave to remain in the UK being targeted for labour exploitation.[1169]

Evidence of risk factors on user-to-user services

9.47 We consider that the risk factors discussed here are likely to increase the risks of harm relating to the human trafficking offence.

[1160] Metropolitan Police, n.d. Modern Slavery. [accessed 25 September 2023].
[1161] Unseen, 2023. Who Cares? A Review of Reports of Exploitation in the Care Sector. [accessed 17 October 2024].
[1162] IOM, 2022. Migrants and their vulnerability. [accessed 17 October 2024].
[1163] Europol, 2024. 13 victims of human trafficking safeguarded in Spain. [accessed 17 October 2024].
[1164] Council of Europe, 2022. Online and technology-facilitated trafficking in human beings: Full report. [accessed 17 October 2024].
[1165] EU Monitor, 2023. Explanatory Memorandum. [accessed 12 August 2023].
[1166] Europol, 2024. Tackling threats, addressing challenges - Europol's response to migrant smuggling and trafficking in human beings in 2023 and onwards. [accessed 17 October 2024].
[1167] Home Office, 2024. Irregular migration to the UK, year ending June 2024. [accessed 17 October 2024].
[1168] Justice and Care, 2023. Modern Slavery Issue Brief: Addressing vulnerability to modern slavery in a growing tide of migration. [accessed 17 October 2024].
[1169] Programme Challenger, 2024. Greater Manchester guide to exploitation in the care sector. [accessed 17 October 2024].

Risk factors: Service types

9.48 Research indicates that social media services, private messaging services and marketplace and listings services online user-to-user pornography services can be used to commit or facilitate offences related to human trafficking offence.

Social media services and messaging services

9.49 Social media and messaging services provide means for perpetrators to communicate with victims and take actions that facilitate their exploitation. Depending on the context they may use social media to maintain or create a relationship with someone before exploitation, with perpetrators "*actively 'hunting' those who they deem as vulnerable to falling victim to trafficking, or passively 'fishing' for potential victims by posting advertisements and waiting for potential victims to respond*."[1170]

9.50 Traffickers can use social media services to gain an insight into people's lives by leveraging the information they share to "*exploit vulnerabilities and tailor escalating manipulation tactics"*.[1171] Victims can then be led to other services such as job boards that have fake advertisements, or private messaging services.[1172] A US report says that traffickers use technology to increase the efficiency of their operations and use social media services to gain insight into people's lives.[1173]

9.51 Social media can be used by traffickers to advertise supposed job opportunities or ways to make money which are used to trap victims into trafficking.[1174] Crest Advisory, in a report looking at the onset of county lines activity, described how certain social media channels served a 'broadcast' function, glamourising a certain lifestyle ostensibly funded by drug dealing. Those engaging with these social media channels were then being signposted to private messaging chats with end-to-end encryption.[1175] Research by the Alliance to Counter Crime Online suggests that perpetrators take advantage of services which offer encryption and anonymisation technology to help them carry out human trafficking offence.[1176]

9.52 Further, the London Rescue and Response County Lines Project (R&R) identifies that social media is a primary facilitator of grooming for, and recruitment into, illicit drug enterprises.[1177]

[1170] UNODC, 2020. Global Report on Trafficking in Persons 2020: Chapter 5. [accessed 17 October 2024].
[1171] Administration for children and families (Contreras, J. and Chon, K.), 2022. Technology's complicated relationship with human trafficking. [accessed 25 September 2023].
[1172] Administration for children and families (Contreras, J. and Chon, K.), 2022.
[1173] "*In 2020, researchers identified a 125% year-on-year increase in the number of reports of trafficking recruitment on Facebook, and a 95% increase in similar reports on Instagram. Individuals often share posts, updates and content that describe their hobbies and interests and express their frustrations and hardships. Traffickers leverage this information to exploit people's vulnerabilities and develop tactics to escalate manipulation, grooming individuals by offering empathy and support, forming emotional connections, and building trust and confidence. In cases of labour exploitation, traffickers will use social media to scout job seekers or those experiencing financial hardships and then use online job boards and employment websites to recruit them through false advertisements"*. Source: Administration for children and families (Contreras, J. and Chon, K.), 2022.
[1174] Centre for Social Justice, 2024. Criminal exploitation. [accessed 17 October 2024].
[1175] Crest (Caluori, J, Mooney, B. and Kirk, E.), 2022. Running out of credit: Mobile phone tech and the birth of county lines. [accessed 25 September 2023].
[1176] Alliance to Counter Crime Online, n.d. Human trafficking: How Social media Fuels Modern Day Slavery. [accessed 25 September 2023].
[1177] Rescue and response, 2019. Rescue and response county lines project. [accessed 12 May 2023].

Marketplaces and listings services

9.53 Online marketplaces are identified as a common type of service used by traffickers in reports from the Council of Europe and from the United Nations Office on Drugs and Crime. These services provide potential perpetrators with an opportunity to post false or misleading opportunities to recruit their targets and usually provide the means for initiating communication.[1178] [1179]

9.54 Based on their work supporting victims and survivors, the Children's Society have noted marketplace and listings services being used to facilitate child criminal exploitation. They describe examples of children in need of money being targeted via these services, as well as perpetrators using the service to gain access to a property.[1180]

9.55 As job adverts are frequently discussed in evidence about human trafficking, and recruitment services are an online space featuring many job adverts, we consider listings services focusing on recruitment and employment to be a relevant risk factor.

9.56 As noted above, there is evidence that adverts for the services of sex workers can feature victims of human trafficking.[1181] As sex workers' services can be advertised on marketplaces and/or listings platforms, we consider these services to also be a risk factor for human trafficking.

9.57 Technological developments over the past decade have created new opportunities for sexual exploitation. Research in this area has described how Adult Service Websites[1182] (ASWs) "*can* (*knowingly or unknowingly) facilitate modern slavery abuses, as victims of sexual exploitation are advertised alongside sex workers.*"[1183] ASWs often use listings as a way to advertise the services of sex workers.

9.58 An inquiry from the Scottish Cross-Party Group on Commercial Sexual Exploitation found many cases of sexual exploitation had taken place via an adult service website.[1184] In an NCA report on modern slavery, they found that online adult services may unwittingly play a role in *"expanding offenders' client bases"*.[1185]

User-to-user pornography services

9.59 User-to-user pornography services were also found to be a risk factor in human trafficking offence, particularly when used for the purpose of sexual exploitation and abuse. Evidence shows that there are examples of pornography on services where the content or services

[1178] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons 2020. [accessed 5 July 2023].
[1179] Council of Europe, 2022. Online and technology-facilitated trafficking in human beings: Full report. [accessed 17 October 2024].
[1180] Children's Society response to the Protection of Children May 2024 Consultation. pp. 8-9.
[1181] All-Party Parliamentary Group on Commercial Sexual Exploitation, 2023. Pornography Regulation: the case for Parliamentary Reform; and All-Party Parliamentary Group on Commercial Sexual Exploitation, 2021. Bust the Business Model: How to stop sex trafficking and sexual exploitation in the UK. [both accessed 17 October 2024].
[1182] As noted in the Illegal Content Judgement Guidance, ASWs need not necessarily be a website, it can also be an app, forum or another service type.
[1183] Modern Slavery & Human Rights, 2023. The role of adult service websites in addressing modern slavery. [accessed 13 August 2024].
[1184] Cross-Party Group on Commercial Sexual Exploitation, 2021. Online Pimping: An inquiry into Sexual Exploitation Advertising Websites. [accessed 17 October 2024].
[1185] NCA, n.d. Modern slavery and human trafficking. [accessed 25 September 2023].

feature victims of trafficking.[1186] There are also examples of CSAM being hosted on user-to-user pornography services featuring trafficking victims.[1187]

9.60 There is also a risk that a user-to-user pornography service can feature adverts and listings for sex workers' services, that could include services of trafficking victims forced into sex work.[1188]

Gaming services

9.61 There is evidence indicating that victims or survivors can be targeted via gaming services.[1189] Charities such as Catch-22 and the Children's Society have supported child victims/survivors of criminal exploitation who have been recruited into trafficking via online gaming. In this context, perpetrators may take advantage of how communication and interaction between strangers is integral to how some children play games online. There are also examples of children being recruited and groomed for exploitation after being offered gaming credits.[1190]

Risk factors: User base

User base size

9.62 Across our evidence base, services with a large user base are frequently cited as being used to commit or facilitate human trafficking offence.

9.63 Traffickers will in some cases target particularly vulnerable people who have shared their stories on social media services. As mentioned above, "*traffickers leverage this information to exploit vulnerabilities and tailor escalating manipulation tactics, grooming individuals by offering empathy and support, forming emotional connections, and building trust and confidence*".[1191]

User base demographics

9.64 The following section outlines key evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

[1186] All-Party Parliamentary Group on Commercial Sexual Exploitation, 2023. Pornography Regulation: the case for Parliamentary Reform [accessed 17 October 2024].
[1187] BBC News, 2021. Pornhub owner settles with Girls Do Porn victims over videos, BBC News 19 October. [accessed 17 October 2024].
[1188] All-Party Parliamentary Group on Commercial Sexual Exploitation, 2021. Bust the Business Model: How to stop sex trafficking and sexual exploitation in the UK. [accessed 17 October 2024]; and UNODC, 2020. Global Report on Trafficking in Persons 2020: Chapter 5. [accessed 17 October 2024].
[1189] Children and Young People's Centre for Justice, 2023. Understanding Child Criminal Exploitation in Scotland: A Scoping Review. [accessed 1st November 2024]. Council of Europe, 2022. Online and technology-facilitated trafficking in human beings: Full report. [accessed 17 October 2024].
[1190] UK Parliament, 2023. Oral evidence submitted by the NCA. [accessed 17 October 2024]; Children's Society response to the Protection of Children May 2024 Consultation.; Mararike, S., (2021) *Dealers are using Fortnite treats to groom children as drug mules*, The Sunday Times, 21 March. [accessed 17 October 2024].
[1191] Administration for children and families (Contreras, J. and Chon, K.), 2022. Technology's complicated relationship with human trafficking. [accessed 25 September 2023].

9.65 The UK-based charity Unseen[1192] does not identify a typical victim or survivor of modern slavery offences, but cites "*poverty, lack of education, unstable social and political conditions, economic imbalances and war*" as key factors that contribute to a person's vulnerability to becoming a victim.[1193] Those who post personal information about financial hardships, or their struggles with self-esteem, family or anxiety, are also among the groups that could be targeted.[1194]

9.66 A report about children, young adults and modern slavery identifies children as more vulnerable to exploitation "*simply because of their age, experience, knowledge and maturity level."* They note other risk factors among children include experiences of the care system, poverty, unstable immigration status, and special educational needs.[1195]

9.67 Data suggests that user base demographics including **socio-economic factors** and **mental health** could lead to an increased risk of harm to individuals. The evidence also suggests that certain types of individuals may be targeted by perpetrators depending on the purpose and nature of trafficking. Those looking to exploit people looking to migrate to the UK may target victims based on nationality. Perpetrators of county lines may target children. Those looking to commit trafficking for the purpose of sexual exploitation may target individuals based on age and gender.

Risk factors: Functionalities and recommender systems

User identification

User profiles and fake user profiles

9.68 As with the CSEA offences presented in the Child sexual exploitation and abuse chapter, user profiles are also used to exploit others by acting as a tool to identify vulnerable people. A United Nations report spoke of traffickers *'hunting*' and explains how user profiles are used by traffickers to hunt both for victims and for potential buyers of exploitative services.[1196] The targets of this approach are not random but chosen for specific characteristics that are presumably identified using victims' and survivors' user profiles.

9.69 Fake user profiles can also be used by traffickers to hide their genuine identities to manipulate others.[1197] An example was provided where a trafficker used two profiles representing fake identities; one to message abusive content to their victims and the other to be understanding of their situation. This encouraged victims to trust the perpetrator.[1198]

[1192] Unseen is a UK charity which provides safehouses and support in the community for survivors of trafficking and modern slavery.
[1193] Unseen, n.d. Modern Slavery Facts and Figures. [accessed 25 September 2023].
[1194] Alliance to Counter Crime Online, n.d. Human trafficking: How Social media Fuels Modern Day Slavery. [accessed 25 September 2023].
[1195] The Rights Lab and ECPAT UK (Every Child Protected Against Trafficking), 2024. Prevention and identification of children and young adults experiencing, or at risk of, modern slavery in the UK. [accessed 17 October 2024].
[1196] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons. [accessed 25 September 2023].
[1197] Di Nicola, A., Baratto, G., and Martini, E. 2017. Surf and Sound: The role of the Internet in People Smuggling and Human Trafficking. [accessed 13 November 2024].
[1198] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons. [accessed 25 September 2023].

User communications

Ephemeral messaging

9.70 Studies suggest that county lines exploitation is facilitated through service applications which offer ephemeral messaging. The ability for a message to disappear is identified as making it harder to identify perpetrators' activities.[1199]

9.71 The Children's Society note that in their experience of supporting victims of child criminal exploitation, they have seen examples of online services being used that "*leave little trace as messages are deleted immediately*", indicating the possible use of ephemeral messaging functionalities to recruit victims.[1200]

Posting content (images, videos, hashtags, emojis)[1201] and livestreaming

9.72 Traffickers using online services to facilitate exploitation often post adverts or opportunities to engage with victims. In some cases of criminal exploitation, posting content can help a trafficker not only state there is an opportunity for financial reward, but also to more widely glamourise and promote the supposed lifestyle that the target could enjoy.[1202]

9.73 As human trafficking offence can involve sexual exploitation of adults, content can also be posted to attract the attention of buyers. This can involve using public posts with "*coded signals to communicate specific information*" about what is for sale.[1203]

9.74 A report from the Alliance to Counter Crime Online describes how *"traffickers will advertise their victims online, using a blend of emojis and coded images to indicate that people are for sale."* [1204] An investigation from 2019 found potential large-scale modern slavery being facilitated online via hashtags being used for sale of domestic workers.[1205]

9.75 The Council of Europe also reports cases where sexual exploitation and abuse of adults and children has been livestreamed.[1206]

Direct messaging and encrypted messaging

9.76 Direct messaging and encrypted messaging can allow traffickers to communicate with one another and with victims. There is evidence that these forms of communication have been used to facilitate child criminal exploitation.[1207] They can also be used in cases of sexual

---

[1199] Crest (Caluori, J, Mooney, B. and Kirk, E.), 2022. Running out of credit: Mobile phone tech and the birth of county lines. [accessed 25 September 2023]. Children and Young People's Centre for Justice, 2023. Understanding Child Criminal Exploitation in Scotland: A Scoping Review. [accessed 1st November 2024].

[1200] Children's Society response to the Protection of Children May 2024 Consultation.

[1201] Where the evidence below makes reference to 'advertising,' this is typically used to refer to content which has the effect of marketing or promoting goods and services, rather than paid advertising, which is detailed in the Revenue Model section below.

[1202] The Children's Society, 2019. Counting lives: Responding to children who are criminally exploited. [accessed 13 November 2024].

[1203] Human Trafficking Front, 2023. Social media and child sex trafficking. [accessed 17 October 2024].

[1204] Alliance to Counter Crime Online, n.d. Human trafficking: How Social media Fuels Modern Day Slavery. [accessed 25 September 2023].

[1205] Although one of the hashtags involved was subsequently banned by the service provider, the potential for using coded language in hashtags means we consider the risk related to this feature remains. Source: BBC News, 2019. Technology and human trafficking. [accessed 17 October 2024].

[1206] Council of Europe, 2022. Online and technology-facilitated trafficking in human beings: Full report. [accessed 17 October 2024].

[1207] Children and Young People's Centre for Justice, 2023. Understanding Child Criminal Exploitation in Scotland: A Scoping Review. [accessed 1st November 2024].

exploitation for communication between traffickers and buyers of sexual services carried out by the person that is being trafficked. After initially advertising an opportunity, direct and/or encrypted messaging is often used to continue communication between a perpetrator and their target as a way of facilitating exploitation, such as providing information about and discussing a supposed job opportunity. One report shares that traffickers have in some cases continued communicating with victims/survivors via encrypted communication after exploitation has ended, to try and intimidate and dissuade them from seeking justice.[1208]

Posting or sending location information

9.77 Evidence indicates that functionalities involving location information can help facilitate human trafficking offence, as this enhances the ability for perpetrators to remotely 'mother' a victim.[1209] There is evidence of child targets of trafficking and exploitation having their location monitored via services with location functionalities.[1210] Another report describes how location information could be used to identify potential victims in a specific geographical area, for example on a dating app, as well as to track a victim/survivor's location, even after their exploitation has seemingly ended.[1211]

User networking

User groups and user connections

9.78 User groups are a risk factor for human trafficking offence, given how they can be used in a range of ways such as targeting victims and facilitating exploitation and abuse. The Council of Europe notes that user groups made up of individuals of the same nationality looking for employment, with the intention of providing mutual support, are spaces in which false advertisements are circulated.[1212] In addition, during the trial of Dominique Pélicot in France in 2024, there was evidence that he recruited men to take part in the sexual exploitation and abuse of his wife via an online forum.[1213]

9.79 As previously noted, perpetrators may have pre-existing relationships with potential victims via friends and family.[1214] In this context we consider that services with user connections are a risk factor for human trafficking offence, as they provide perpetrators with ways of contacting victims/survivors and facilitating their exploitation, through direct or mutual contacts.

[1208] Council of Europe, 2022. Online and technology-facilitated trafficking in human beings: Full report. [accessed 17 October 2024].
[1209] Children's Society response to the Protection of Children May 2024 Consultation.
[1210] The Children's Society, 2019. Counting lives: Responding to children who are criminally exploited. [accessed 13 November 2024].
[1211] Polaris Project, 2018. A roadmap for systems and industries to prevent and disrupt human trafficking: Social media. [accessed 17 October 2024].
[1212] This report also notes communication between traffickers and victims within closed online groups, but it is unclear if these refer to groups with more than two people or not. Source: Council of Europe, 2022. Online and technology-facilitated trafficking in human beings: Full report. [accessed 17 October 2024].
[1213] Robins-Early, N. (The Guardian). 2024. The anonymous, anything-goes forum at the heart of the Pelicot rape case, 12 October. [accessed 28 October 2024].
[1214] For example, government guidance about county lines notes that in cases of criminal exploitation victims sometimes have pre-existing connections with perpetrators. Source: Home Office, 2023. Criminal exploitation of children, young people and vulnerable adults: county lines. [accessed 17 October 2024].

9.80 There is evidence that user groups and connections are a risk factor for child criminal exploitation. User groups have been used to share 'adverts' for participation in criminal activity, and connections can be leveraged by gangs as a way of encouraging targets to be associated with them if they are seen as having high status.[1215]

Transactions and offers

Posting goods or services for sale

9.81 The ability to post goods and services for sale can be used in human trafficking offence. This is particularly prevalent on adult service websites.[1216] In a study of advertisements on the "*largest dedicated platform for sex work services in the UK,*" the majority of advertisements (58.3%) were found to include at least one indicator of human trafficking. For more details see the chapter on sexual exploitation of adults.

9.82 A United Nations report indicates that advertisements are used as tools by exploiters to recruit victims. The report also specifies that *"online marketplace sites, on which anyone can post or browse advertisements to sell or buy any service (from job vacancies to the sale of equipment, cars and clothes), are being used to advertise services obtained from victims of human trafficking"*. This suggests that the 'advertisements' described in the report can include posts that have the effect of advertising or promoting services by posting them for sale. They can also be used in recruitment by allowing traffickers to make 'fake job advertisements' that offer well-paid jobs or include wording that describes the possibility of living a luxurious life or getting jobs in industries such as modelling or entertainment.[1217]

Risk factors: Business model and commercial profile

Revenue models

9.83 Services that generate revenue through advertising can be at risk, as offenders may be able to use adverts to lure victims, who can then be coerced or controlled into sexual activities.

9.84 As noted in the chapter on Sexual exploitation of adults, services that generate revenue through advertising may enable perpetrators to reach potential targets of trafficking for sexual exploitation and abuse, as well as customers, in cases of forced commercial sexual exploitation.

9.85 A report from the United Nations Office on Drugs and Crime[1218] set out that services which provide classified listings or other advertising provide sexual traffickers with an effective mechanism to place attractive but false career advertisements to recruit and entrap victims.

9.86 In addition, a study, 'The Role of Technology in Domestic Minor Sex Trafficking' recognised that some services which offer online advertising have been *"profiting from prostitution ads*".[1219] We think that a significant proportion of this is likely to have been presented by trafficked sex workers.

[1215] Rescue and Response, 2020. Year 2 Strategic Assessment. [accessed 1st November 2024].
[1216] Modern Slavery & Human Rights, n.d. The role of adult service websites in addressing modern slavery. [accessed 25 September 2023].
[1217] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons. [accessed 25 September 2023].
[1218] United Nations Office on Drugs and Crime, 2020..
[1219] Thorn, 2018. Survivor Insights: The role of technology in domestic minor sex trafficking, p38 [accessed 5 July 2023].

9.87 Services that generate revenue through the sale of online gifts or tokens may also be at greater risk of being exploited by perpetrators of online grooming. Accounts from journalists as well as charities supporting child victims/survivors of child criminal exploitation note examples of children being recruited and groomed for exploitation after being offered gaming credits.[1220]

[1220] Children's Society response to the Protection of Children May 2024 Consultation.; Mararike, S., 2021. Dealers are using Fortnite treats to groom children as drug mules, The Sunday Times, 21 March. [accessed 17 October 2024].

10. Unlawful immigration

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for unlawful immigration offences: how harm manifests online, and risk factors

This chapter covers offences relating to unlawful immigration.

People entering the UK without permission may do this for a range of reasons, which can include fleeing another country owing to fear of physical harm or forms of illegitimate persecution. This chapter talks about the risks of harm from unlawful immigration, which often result from actions taken by people smugglers. These can be widespread, including trauma, financial hardship, injury or even death for those undertaking potentially dangerous routes of entry. Evidence also shows that some people smugglers also commit human trafficking offences, meaning that some irregular migrants face risks of harm related to those offences.

Service type risk factors:

Social media services and **messaging services** are shown to be risk factors for unlawful immigration offences. Among other elements, they are used to target people and to advertise services. Often, perpetrators will identify someone on social media before transitioning to an encrypted private messaging service to further the facilitation of the offence.

Functionalities and recommender systems risk factors:

User profiles can be exploited by a perpetrator looking to build trust with their targets, by providing a sense of legitimacy to their enterprise. Closed **user groups** were found to be a risk factor, with smugglers using these groups to share information which could help facilitate the offences.

Posting content is also used to promote illegal services. Posting content could also be used by a smuggler to build trust in the journey's safety.

Direct messaging and **encrypted messaging** can allow smugglers to communicate with people using their services and others facilitating unlawful migration.

Introduction

10.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to unlawful immigration offences listed under 'Relevant offences'; and
- The use of these services for the commission and/or facilitation of these offences (collectively, the 'risks of harm').

10.2 We set out the characteristics of User-to-user (U2U) services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the effect of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

10.3 The evidence presented below focuses on unlawful immigration offences on U2U services with reference to the concept of 'people smuggling' to describe some offences.[1221] It is important to recognise that these are harms and offences which are also experienced by children. Other, related offences are discussed in other chapters, see in particular 'Human trafficking'.

10.4 Ofcom has reviewed the available evidence to develop this assessment of the risks of unlawful immigration and how they manifest on U2U services. It may not be an exhaustive account of the possible uses of online services to facilitate these offences.

Relevant offences

10.5 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding unlawful immigration, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act:

a) Illegal entry and similar offences[1222]
b) Assisting unlawful immigration[1223]

10.6 The Act also covers inchoate offences such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and, in relation to offences in Scotland, being involved art and part in the commission of these offences).

10.7 The **unlawful immigration offences** covered in this chapter are 'illegal entry and similar offences' and 'unlawful immigration' (listed in (a) and (b) above). 'Illegal entry' means an individual entering the United Kingdom in breach of a deportation order, entering without permission to remain,[1224] or without entry clearance when the individual needs it. A person commits the offence of 'unlawful immigration' if they do an act which facilitates a breach or attempted breach of immigration law by an individual who is not a national of the United Kingdom – and where they know or have reasonable cause for believing this to be the case. It is usually encouraged by organised crime. This is distinct from other, legal, means of immigration, which are not covered in this chapter. Online aspects of unlawful immigration could include the sale of counterfeit travel documents such as passports, visas and identification papers, as well as the sale of crossings.[1225]

[1221] 'People smuggling' and 'people trafficking' are different concepts in law. Offences relating to 'people smuggling' will generally relate to the Immigration Act offences, whereas 'people trafficking' will generally be offences under the Modern Slavery Act, Crown Prosecution Service, Updated 6 July 2022. Source: Crown Prosecution Service, 2022. Modern Slavery, Human Trafficking and Smuggling. [accessed 25 September 2023].
[1222] Section 24(A1), (B1), (C1) or (D1) of the Immigration Act 1971.
[1223] Section 25 of the Immigration Act 1971.
[1224] Generally referred to as 'leave to enter or remain'. Source: Home Office, 2023. Immigration Rules. [accessed 25 September 2023].
[1225] See the Fraud and financial services chapter for information about other offences regarding the sale of counterfeit products or other fraud involving mis-selling.

10.8 For more details on the offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance or ICJG.

How unlawful immigration offences manifest online

10.9 This section is an overview which looks at how unlawful immigration offences manifest online, and how individuals may be at risk of harm.

10.10 It is not possible to determine the total number of migrants who have come to the UK unlawfully.[1226] Nonetheless, government statistics report an increase in irregular[1227] migration over the past few years due to people arriving across the English Channel in small boats, the predominant recorded method since 2020. Between July 2023 and June 2024, 38,784 people were reportedly detected arriving.[1228]

10.11 It is worth noting that there are legal routes for entering the UK with the intention of claiming asylum, but also that many of those arriving by irregular means are granted refugee status or other leave.[1229] The research and guidance we must provide in our consultation refers to unlawful entry and the assistance of unlawful entry only; that is, entering the UK without permission ('leave') to do so.

10.12 Evidence that looks at why individuals attempt to irregularly migrate to the UK highlights a variety of causes. Asylum seekers are often motivated by geopolitical factors in their countries of origin, such as ongoing war or conflict, and a desire for safety from harm.[1230] Smugglers can take advantage of these needs by selling false documents and offering transportation services. They can exert a large influence on individual's migration journeys, for example, determining the method of transport, or, in some cases, selecting a destination country.[1231]

10.13 Evidence relating to the EU, highly relevant to the UK context, suggests that more than 90% of irregular migrants coming to the EU "*make use of the services of smugglers, mostly organised in criminal groups."* They note that many smugglers are also involved in other criminal activity such as human trafficking and smuggling of illegal drugs and weapons.[1232] The National Crime Agency (NCA) have similarly argued that many human trafficking cases also involve organised immigration crime (OIC).[1233]

[1226] There are many reasons for this, including the sometimes-clandestine nature of irregular migration. On the use of 'irregular migration', terminology around unlawful immigration is highly contentious. The UK government reports statistics based on 'irregular migration' rather than 'unlawful immigration', and we have adopted that wording in this chapter.
[1227] "Irregular" is the term used by the UK government to refer to migration that occurs outside of the law. "Undocumented" migration is also used.
[1228] Statistics on small boats include individuals who were detected on arrival to the UK or detected in the Channel and subsequently brought to the UK. It is also worth noting the latest figures show a 26% decrease in irregular arrivals compared to the previous year. Source: Home Office, 2024. Irregular migration to the UK, year ending June 2024. [accessed 5 September 2024].
[1229] Between January 2018 and the end of June 2024, 35,396 asylum applicants who had arrived on a small boat had been granted refugee status or leave; this was the most common outcome of applications during this period. Source: Home Office, 2024. Irregular migration to the UK, year ending June 2024. [accessed 5 September 2024].
[1230] The Migration Observatory, 2024. UK policies to deter people from claiming asylum. [accessed 8 August 2024].
[1231] Refugee Council (Crawley, H.), 2010. Chance or choice? Understanding why asylum seekers come to the UK. [accessed 8 August 2024].
[1232] EU Monitor, 2023. Explanatory Memorandum. [accessed 12th August 2023].
[1233] National Crime Agency (NCA) (Arnold, P.), 2024. Written evidence to the Modern Slavery Act 2015 Committee. [accessed 14 November 2024].

10.14 There is evidence to suggest that online spaces are increasingly being used by organised criminals to facilitate the travel of migrants, often facilitated by organised immigration crime. The National Crime Agency (NCA) have said that "*Social media is a key component of the OIC facilitators' business model. Facilitators use social media proactively to advertise migration services and target potential customers*."[1234]

10.15 The NCA has worked with social media services to take down illegal content relating to OIC, via a collaboration with X (formerly Twitter), TikTok, Meta and YouTube.[1235] In August 2023, the Home Office announced that this collaboration with social media services would be extended, with new funding increasing the capacity and capability of law enforcement to identify this content online.[1236]

10.16 The NCA explained in 2024 that the common types of illegal content relating to these offences on social media services are smuggling services, the supply of false documents, a combination of both of these, and assistance with "*fraudulent services (for example, visa or asylum abuse)*."[1237] Indicators of such content can include advertisements for documents or visas, directing enquiries to private channels for communication away from public-facing social media, and videos containing tutorial content on how to cross borders illegally.[1238]

10.17 A report by Europol in 2021 further highlighted the increased role that digital technologies play in migrant smuggling. The report says that migrant smugglers have expanded their use of social media services and mobile applications to offer illegal services.[1239]

10.18 Due to the nature of unlawful immigration offences, much of the content relating to these will likely be aimed at users outside of the UK. However, it is worth noting that there are also examples of smugglers illegally transporting or attempting to transport people out of the UK – meaning that offences can be relevant to people residing in the UK as well.[1240]

Risks of harm to individuals presented by unlawful immigration offences

10.19 As well as the risks they might face in their countries of origin, there are multiple risks associated with people entering the UK through irregular routes. These include, but are not limited to, the risk of trauma occurring from undertaking a potentially perilous journey,[1241] the high cost of smugglers' services for individuals who may have few financial resources to draw on,[1242] and the related risk of being defrauded by someone who takes payment but does not provide any transport services. Victims usually have limited means to report this

[1234] NCA, 2024.
[1235] NCA, 2024. 12,000 takedowns as NCA leads blitz on people smugglers' social media accounts. 24 July 2024. [accessed 8 August 2024].
[1236] Home Office, 2023. New tech partnerships to stop the boats. [accessed 25 September 2023].
[1237] NCA, 2024.
[1238] NCA, 2024.
[1239] The report states that these services are *"frequently used for various purposes such as advertising, recruitment, communication, coordination, guidance, money transfer or monitoring law enforcement activities"* Europol, 2022. European Migrant Smuggling Centre - 6th Annual Report. [accessed 25 September 2023].
[1240] NCA, 2024. Pair jailed for attempt to smuggle migrants out of the UK. [accessed 12 August 2024].
[1241] For example, see World Health Organization, 2023. Mental health of refugees and migrants: risk and protective factors and access to care. [accessed 12 August 2023].
[1242] Hymas, C, 2024. Migrant smugglers quadruple price of passage across the Channel, The Telegraph, 20 May [accessed 8 August 2024].

to relevant authorities.[1243] Financial costs can leave people vulnerable to exploitation (see below) and debt bondage.[1244]

10.20 There is also the risk of death or serious injury. Recent data estimates that at least 274 migrants have died attempting to migrate to the UK during January 2014 to September 2024. Those who died include adults and children.[1245] In addition, smugglers have been known to use violence against irregular migrants during transportation as well as to extract payment.[1246]

10.21 People coming to the UK by irregular means are also at heightened risk of being victims of other harms and offences, some we discuss in this chapter, such as human trafficking and sexual exploitation.[1247] Organised criminals may be involved in both smuggling and human trafficking, meaning irregular migrants can be targeted for trafficking. The expensive fees of smugglers can lead to debt bondage, with payment being taken via sexual or labour exploitation.[1248]

10.22 Relatedly, the financial proceeds of people smuggling can be used to fund other criminal activities, such as terrorism.[1249]

Evidence of risk factors on user-to-user services

10.23 We consider that the risk factors below are likely to increase the risks of harm relating to unlawful immigration offences.

Risk factors: Service types

10.24 Research indicates that social media services and private messaging services can be used to commit or facilitate offences related to unlawful immigration offences.

Social media services and messaging services

10.25 Social media services are a risk factor for unlawful immigration offences.[1250] Between November 2021 and June 2024, the NCA's collaboration with social media services led to the suspension of nearly 12,000 posts, pages or accounts.[1251] It has been noted that social

[1243] This report provides evidence of online posts that expose fraudulent smugglers. Diba, P., Papanicolaou, G. & Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK. [accessed 25 September 2023].
[1244] NCA, 2024.
[1245] Data shows that of the 263 who died or went missing travelling the English Channel to the UK during this time, 248 had died due to reasons including hazardous transport and drowning. Missing Migrants Project, 2024. Migration Within Europe, Latest Incidents. [accessed 25th October 2024].
[1246] Europol, 2023. Criminal Networks in Migrant Smuggling. [accessed 12th August 2024].
[1247] For more information about these kinds of illegal harm see the relevant register of Risks chapters (Human trafficking, Sexual exploitation of adults and Child sexual abuse and exploitation).
[1248] Europol, 2023.
[1249] Europol, 2024. 21 arrested in hit against migrant smuggling across the EU-Russian border. [accessed 10 October 2024].
[1250] Infomigrants is co-financed by the EU and in partnership with three major European media sources: France Médias Monde (French), Deutsche Welle (German) and ANSA (Italian). Source: Infomigrants (Alboz, D.), 2016. Social media networks, the best friend of smugglers. [accessed 25 September 2023].
[1251] NCA, 2024. 12,000 takedowns as NCA leads blitz on people smugglers' social media accounts. 24 July 2024. [accessed 8 August 2024].

media services, particularly those with encryption, provide a *"really good, dynamic, agile way for people to move migrants between them, and for groups to communicate"*.[1252]

10.26 Evidence suggests that smugglers will then encourage their target to move to messaging services, usually with encryption.[1253] [1254] Smugglers have been noted to share guidance about how to cross borders using such services.[1255] The NCA have also described smugglers being contacted by individuals "*making contact on behalf of a family, community or group of people, extending reach and increasing opportunities or sub-contractors or profiteers to operate*."[1256]

10.27 People smugglers use a range of techniques to advertise their services on social media. Europol have also noted that smuggling networks post 'sophisticated and professional' advertisements on social media to recruit collaborators who can transport migrants.[1257]

10.28 One example of how private messaging and social media are used to facilitate unlawful immigration includes evidence from Europol, who, in 2021, found a total of 455 social media accounts facilitating unlawful immigration from Belarus to Europe. At the time they stated *"the new Belarusian migratory route is heavily advertised to migrants on social media and instant messaging applications, which represents a significant pull factor. The misuse of these online platforms by facilitators [has] led to a large increase of departures and irregular border crossings"*.[1258]

Services enabling users to build online communities

10.29 Service types which encourage community building are at risk of enabling unlawful immigration offences. Evidence suggests that offenders can use online communities to target vulnerable users by advertising unlawful immigration services, and by building trust with those seeking their services.[1259] Also, evidence shows that smugglers use private groups to share information about crossing itineraries and departure points.[1260] This can include information on routes, border closures, transport services and the cost of arranging trips.[1261]

[1252] Gentleman, A., 2020. Social media refuse to pull people-smuggling pages, MPs told, *The Guardian*, 3 September. [accessed 25 September 2023].
[1253] Europol, 2024. Tackling threats, addressing challenges: Europol's response to migrant smuggling and trafficking in human beings in 2023 and onwards. [accessed 12th August 2024].
[1254]Diba, P., Papanicolaou, G. and Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK. [accessed 25 September 2023].
[1255] Europol, 2023. Criminal Networks in Migrant Smuggling. [accessed 12th August 2024].
[1256] NCA, 2024.
[1257] Europol, 2024. Tackling threats, addressing challenges: Europol's response to migrant smuggling and trafficking in human beings in 2023 and onwards. [accessed 12th August 2024].
[1258] Europol, n.d. Europol coordinates referral action targeting migrant smuggling from Belarus. [accessed 25 September 2023]. Whilst this particular illegal immigration trend may no longer be taking place, it is an example of how these online functionalities can be used to facilitate similar activities in other contexts.
[1259] FATF, 2022. Money Laundering and Terrorist Financing Risks Arising from Migrant Smuggling. [accessed 10 October 2023].
[1260] Infomigrants (Alboz, D.), 2016. Social media networks, the best friend of smugglers. [accessed 25 September 2023].
[1261] Diba, P., Papanicolaou, G. and Antonopoulos, G.A., 2019.

Risk factors: User base

User base size

10.30 Across our evidence base, services with a large user base are frequently cited as being used to commit or facilitate unlawful immigration offences. Integral to this is the popularity of these services in the countries from which migrants are looking to enter the UK.

10.31 There is also evidence to suggest that smugglers use services with a large user base to target individuals. The NCA have described how content can "*utilise legitimate hashtags which attract a large following to capitalise*." They also noted that one video advertising a migration journey "*was viewed by over 1.6 million people*."[1262]

User base demographics

10.32 The following section outlines primary evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

10.33 Irregular migrants come from a wide variety of countries and, consequently, migration that is facilitated online takes place in a variety of languages. Data suggests that user base demographics including **socio-economic factors** and **mental health** could lead to an increased risk of harm to individuals, given that many people who are smuggled face financial hardship and have had adverse life experiences that may have led to mental health challenges.

10.34 The NCA describe how some adverts often target people in desperate circumstances, stating with urgency that those interested should travel before potential changes to immigration measures are made, and providing likely false reassurance about the safety of the method of transport. [1263]

10.35 They also noted that perpetrators can tailor the language of advertisements for irregular migration to "capitalise on individual circumstances, catching the attention of those who are, for example, illegally in the UK and wishing to leave the UK clandestinely."[1264]

Risk factors: Functionalities and recommender systems

User identification

User profiles, fake user profiles, and anonymous user profiles

10.36 User profiles can be exploited to commit or facilitate unlawful immigration offences and can be used by perpetrators to build trust with victims and potential buyers. This is because a user profile can demonstrate trustworthiness, for instance, by displaying an identity that appears real to others when they are looking for information about routes or checking the legitimacy of smugglers. A report discusses how some users who have made illegal

[1262] NCA, 2024.
[1263] NCA, 2024.
[1264] NCA, 2024.

crossings are then contacted on social media to ask details about which routes to take and how to approach an illegal crossing.[1265]

10.37 Further, organised immigration crime facilitators often advertise services by using overt display names and including information in the bio section of an account or a profile.[1266]

User communications

Posting content (videos, emojis, hashtags)

10.38 The posting of content that shares experiences, including information about the route (such as where to stay) and the legitimacy of specific smugglers, can be crucial information in facilitating unlawful immigration offences.

10.39 Posted content can allow potential perpetrators to promote and market their services to attract potential customers and build trust.[1267]

10.40 This content is typically posted on social media services with images of the transport used and the duration of the journey.[1268] Research conducted by BIRN (the Balkan Investigative Reporting Network) found many posts promoting or 'advertising' prices, transport used, and routes proposed, on a large social media service.[1269]

10.41 Videos and images that share common words, emojis, and flags, are often posted to allow the user to track and find content which could help them to enter a country illegally, and to avoid the risk of the content being removed.[1270]

10.42 Videos can be used to explicitly to record a migrant journey, or to persuade people the service is safe.[1271] Sometimes they are used more discreetly, for example by using only the captions to advertise the transport.[1272]

10.43 Smugglers can use hashtags to help advertise their services, whether these are explicit about what services are offered, coded to avoid takedown, or involve unrelated but popular topics (for example, a celebrity) to encourage user engagement.[1273]

Direct messaging and encrypted messaging

10.44 Direct messaging can allow smugglers to communicate with people using their services to attempt irregular migration. Encrypted messaging services are also used by smugglers to communicate both with migrants and with one another when facilitating illegal immigration

---

[1265] Rest of World (Joles, B.), 2022. Inside the risky world of "Migrant TikTok". [accessed 25 September 2023].
[1266] Burgess, S., 2021. Channel deaths: People smugglers touting openly on Facebook, Sky News, 25 November. [accessed 10 October 2024].
[1267] Where the evidence below makes reference to 'advertising,' this is typically used to refer to content which has the effect of marketing or promoting goods and services, rather than paid advertising, which is detailed in the Revenue Model section below.
[1268] Diba, P., Papanicolaou, G. and Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK [accessed 25 September 2023].
[1269] Sinoruka, F., 2022. Rise in TikTok Ads Among Albanians Selling Smuggling Operations to UK. Balkan Insight, 8 August. [accessed 25 September 2023].
[1270] Rest of World (Joles, B.), 2022. Inside the risk world of "Migrant TikTok". [accessed 25 September 2023].
[1271] Kansara, R., Fatima, S. and Dyer, J., 2023. Going undercover to reveal people smugglers' sales tactics, BBC News, 28 October. [accessed 10 October 2024].
[1272] NCA, 2024.
[1273] NCA, 2024.

activities.[1274] [1275] Messaging services are also used for the advertisement of services that smugglers provide.[1276]

User networking

User groups

10.45 User groups are used as spaces to promote unlawful immigration services and share information among smugglers. They are also used to find and establish contact between smugglers and potential migrants.

10.46 An article by InfoMigrants[1277] uncovered closed groups where smugglers share information about crossing itineraries and departure points, and post images of the boats that will be used.

10.47 There is also evidence to suggest that user groups are used by migrants and smugglers to share information about travelling routes, border closures, transport services and the cost of arranging trips.[1278]

10.48 The exchange of goods and services on user groups, such as fake passports and the sale of crossings, can facilitate unlawful immigration offences. This was evidenced in the report *Human Smuggling and the Internet,* which described how researchers found posts that advertised transport services, the sale of counterfeit travel documents such as passports, visas and identification papers on a social media service page. The page also hosted discussions on how to navigate routes into the UK.[1279]

Transactions and offers

Posting goods or services for sale

10.49 There are indications that social media platforms can be used to advertise the services of people smugglers. There are documented examples of people offering services to smuggle individuals to the USA via a social media listings platform.[1280] Europol have also reported that cryptocurrencies are sometimes used when buying migrant smuggling services.[1281]

[1274] Europol, 2024. Tackling threats, addressing challenges: Europol's response to migrant smuggling and trafficking in human beings in 2023 and onwards. [accessed 12th August 2024].
[1275] Gentleman, A, 2020. Social media refuse to pull people-smuggling pages, MPs told, *The Guardian*, 3 September. [accessed 25 September 2023].
[1276] Europol, n.d. Europol coordinates referral action targeting migrant smuggling from Belarus. [accessed 25 September 2023]. Whilst this particular illegal immigration trend may no longer be taking place, it is an example of how these online functionalities can be used to facilitate similar activities in other contexts.
[1277] Infomigrants (Alboz, D.), 2016. Social media networks, the best friend of smugglers. [accessed 25 September 2023].
[1278] Diba, P., Papanicolaou, G. & Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK [accessed 25 September 2023].
[1279] Diba, P., Papanicolaou, G. & Antonopoulos, G.A., 2019.
[1280] Tech Transparency Project, 2022. Facebook Marketplace, WhatsApp Storefronts, and TikTok Videos: How Coyotes Get Creative. [accessed 10 October 2024].
[1281] Europol, 2024. Tackling threats, addressing challenges: Europol's response to migrant smuggling and trafficking in human beings in 2023 and onwards. [accessed 10 October 2024].

11. Fraud and financial services offences

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for fraud and financial offences: how harm manifests online, and risk factors

This chapter covers offences linked to fraud, fraudulent activity and financial services, including the use of misleading statements intended to entice users to engage in relevant investment activity.

Fraud is the most frequently experienced crime in the UK and the risks of harm to individuals from fraud and financial services offences are broad. The 2023 UK Government Fraud Strategy estimated the total economic and social cost of fraud to individuals between 2019 and 2020 to be £6.8bn. Financial loss to victims is not the only result of fraud; the harm can be multi-faceted and can have serious consequences on both mental and physical health.

We have identified the following characteristics of online services that are relevant to risks of harm in relation to these offences.

Service type risk factors:

Our evidence points to fraud taking place on a wide range of services. This can include **social media services**, **messaging services**, **marketplaces and listings services, and dating services**. Victims of fraud can be targeted through one type of service and then potentially be moved to other types of service for further communications and transactions. Different types of fraud may be more common on different types of service; for example, investment or purchase fraud activities are more likely to occur on social media services.

User base risk factors:

Online services with a **large user base** can help fraudsters reach large numbers of potential victims at low cost with minimal effort. In addition, a large user base risks amplifying the initial reach of fraudulent content to an even bigger potential audience via a higher volume of content reactions, posts and re-posts.

Anyone online, of any age, can be a potential victim of fraud; including being targeted for a specific type. **Those who are likely to have the least financial resilience** (for instance, those less able to withstand financial shocks) may fall victim more easily to some specific scams where the supposed gains are promised quickly. Scams include purchase scams offering cheap goods, and loan-fee scams which appear to peak at periods of financial difficulty. In contrast, investment

scams, for example, tend to target **older age groups** with greater disposable income, as well as low-capital individuals attracted by the promise of large returns.

People who have **experienced mental health challenges** are also more likely to have been a victim of an online scam.

Low levels of media literacy may also be a significant factor when assessing the risks of harm. Research suggests that users do not always have the critical skills to recognise fake propositions.

Functionalities and recommender systems risk factors:

Fraud and financial services priority offences can be enabled using a range of functionalities on user-to-user (U2U) services; these are common across most online services and therefore, in principle, almost any service can be attractive to fraudsters.

Creating **fake user profiles** on a U2U service enables fraudsters to commit or facilitate fraud, allowing them to conceal their identity and impersonate legitimate entities such as banks, insurance providers or financial advisors to add legitimacy to false claims. Fraudsters will also make use of people who have many **user connections** to achieve their aims. While **user groups** are used by fraudsters to share knowledge to successfully carry out scams.

Functionalities that allow communications between users can be abused by fraudsters. For example, the ability to communicate via **direct messaging, group messaging or private messaging** may help fraudsters create the appearance of a legitimate organisation when engaging with a user or may help to promote a relationship in a romance scam. Fraudsters can move conversations onto services with **encrypted messaging** to avoid moderation or intervention disrupting their activity, as well as evidence of illegality.

The functionality of **posting goods and services for sale** can enable fraudsters to trick users into paying for goods and services that do not exist or are less valuable than described. **Searching for user-generated content (UGC)** can enable fraudsters or potential fraudsters to find posts offering to supply information, advice and articles (such as stolen bank details) which support the commission of fraud. These functionalities (posting goods and services for sale and searching for UGC) are included in the risk profiles for their role in propagating fraud (see also the 'Proceeds of crime' chapter).

The functionality of **hyperlinking** enables fraudsters to redirect victims to webpages outside of the original service which can then facilitate scams such as purchase scams, advance-fee scams and impersonation scams if the victim shares their personal information or have malicious programmes downloaded onto their device.

The information on **user profiles** can also be used by fraudsters to identify potential victims, such as high net-worth individuals or those who are looking to make

connections, for instance on online dating services. Fraudsters may also join **user groups** to identify potential targets. Users can also encounter fraud in the comments on posts, so **commenting on content** can also be considered a risk factor.

Introduction

11.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to fraud and/or financial services offences listed under 'Relevant offences'; and
- The use of these services for the commission and/or facilitation of these offences (collectively, the 'risks of harm').

11.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind. We also consider the 'financial harm' caused to individuals by fraud.

Relevant offences

11.3 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding fraud and financial services offences, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

11.4 For more details on the offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance or ICJG.

11.5 Offences covered here can be closely related to Proceeds of Crime offences. It is likely in some cases that offences of both kinds might be committed in certain scenarios. We recommend readers familiarise themselves with the Proceeds of crime Register of Risk and ICJG chapters for further context.

Priority offences for fraud

11.6 The priority offences for fraud are:

- **Fraud by false representation:**[1282] It is an offence to 'dishonestly make a false representation' where the person making such a representation intends to make a gain thereby (for themselves or others) or to cause another person loss (or expose them to the risk of loss)
- **Fraud by abuse of position:**[1283] It is an offence to commit fraud by way of a person dishonestly abusing their position

[1282] Section 2 of the Fraud Act 2006.
[1283] Section 4 of the Fraud Act 2006.

- **Making or supplying articles for use in frauds:**[1284] It is an offence to make, adapt, supply or offer to supply any article, knowing that it is designed or adapted for use in the course of or in connection with fraud, or intending that it be used to commit, or assist in the commission of, fraud. In Scotland, this is covered by a separate but similar offence[1285]
- **Participating in fraudulent business carried on by a sole trader:**[1286] It is an offence for a person to knowingly be a party to the carrying-on of a sole trader business with the intent to defraud creditors of any person or for any other fraudulent purpose.

11.7 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and, in relation to offences in Scotland, being involved art and part in the commission of those offences).

Priority offences for financial services

11.8 The priority offences for financial services are:[1287]

- Contravention of the prohibition on carrying on regulated activity in the UK unless authorised or exempt[1288]
- Falsely claiming to be authorised or exempt for the purposes of carrying-on regulated activity[1289] and the contravention of restrictions on financial promotions[1290]
- Making false or misleading statements, or creating false or misleading impressions about relevant investments[1291]

11.9 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and, in relation to offences in Scotland, being involved art and part in the commission of those offences).

How fraud and financial services offences manifest online

11.10 This section is an overview which looks at how fraud and financial services manifest online, and how individuals may be at risks of harm.

Fraud offences: Examples

11.11 There are many types of fraud, and the examples provided here are not exhaustive. They are also likely to change over time. We note that there are varying terms used to describe different types of fraud, and similarly, various ways to consider them. Some examples are:

- **Impersonation fraud**, where fraudsters pretend to be from a legitimate organisation (for example, a financial institution, the NHS, lottery institution, solicitors, government

[1284] Section 7 of the Fraud Act 2006.
[1285] Section 49(3) of Criminal Justice and Licensing (Scotland) Act 2010.
[1286] Section 9 of the Fraud Act 2006.
[1287] These offences are set out in the Financial Services and Markets Act (FMSA) 2000 and the Financial Services Act 2012.
[1288] Section 23 of the FMSA 2000 Act.
[1289] Section 24 of the FMSA 2000 Act.
[1290] Section 25 of the FMSA 2000 Act.
[1291] Section 89 misleading statements or Section 90 misleading impressions of the Financial Services Act 2012.

officials or police officers) and request a payment or information from an individual, potentially via phishing.[1292]

- **Purchase scams**, where a product purchased by a user is not provided, or where provided, is different from the product advertised on the U2U service. For example, sale of fake holidays or counterfeit goods described as genuine, and 'ghost broking'[1293] which involves the sale of fake insurance policies. Fraudsters may pose as known brands.
- **Investment scams**, where fraudsters persuade users to invest in a financial product which does not exist and so a victims' money is stolen. Fraudsters may present themselves as a trustworthy institution, advisor, or someone known to the victim; use pressurising tactics; or promise returns generally not available through mainstream products, for example, by offering cryptocurrency. Pension scams are considered a type of investment scam in this chapter.[1294]
- **Romance scams**, where fraudsters exploit the trust of the victim, who is under the impression that the perpetrator is genuinely interested in building a relationship or friendship. The fraudster typically asks for money or financial information.
- **Employment scams**, in which fake job opportunities are promoted by fraudsters. The fraudsters may ask for payments for processes they say are needed for the victim to secure the role, or to gain personal information from victims.
- **Mule herders**[1295] seeking to recruit **money mules**[1296] may also commit fraud by false representation via user-generated content to trick people into becoming a mule. For example, mule herders may create fake jobs that involve moving money between accounts, including asking the victim to use their own account to help move the money, or to hand over control of their account. Other tactics used to hook potential victims could include a romance scam, where the mule herder exploits the victim's trust to ask them to transfer money or hand over their account details. Users may respond to opportunities to make money, shared via user-generated content, which involve becoming a money mule and earning a commission.

[1292] "*Phishing is when attackers attempt to trick users into doing 'the wrong thing', such as clicking a bad link that will download malware, or direct them to a dodgy website*." National Cyber Security Centre (NCSC), 2018. Phishing attacks: defending your organisation. [accessed 2 October 2023].

[1293] A 'ghost broker' is a term used to describe a fraudster who pretends to be a genuine insurance broker in order to sell fraudulent insurance.

[1294] *"The word 'investment' is used in connection with a wide range of schemes offering income, interest or profit in return for a financial investment. 'Investment' is often used loosely, and sometimes misleadingly, in order to disguise the true nature of a fraud, for example, pyramid schemes, chain letters or other types of scheme where a return depends on persuading others to join. The term 'investment' is commonly used in connection with the purchase of something - such as high value or rare goods, stocks and shares, property, in the expectation that what is purchased will increase in value, and even provide an exceptional return compared to other forms of investment….An investment seminar will hook individuals by offering a return which is more attractive than a conventional investment, and so the return on the outlay is always likely to be exaggerated or unrealistic. It follows that the essential message which applies to other scams applies equally to investments. If it looks too good to be true, it probably is."* Source: Home Office, 2023. Counting rules for recorded crime. [accessed 1 August 2023].

[1295] Mule herders are people who recruit money mules. A money mule is someone who lets criminals use their bank account to move money. See also chapter 6N: Proceeds of Crime offences for more information on money mules.

[1296] A money mule is someone who receives money from a third party in their bank account and transfers it somewhere else, or who withdraws it as cash and gives it to someone else, obtaining a commission for it or payments in kind.

- User-generated content can also be used to supply **stolen identity or banking credentials,**[1297] such as stolen personal information, credit card details, or fraud 'how-to' guides and fake passports.

11.12 Fraud offences can manifest in complex ways; there can be overlaps of the above types of fraud. For example, although we identified impersonation fraud specifically, impersonation can be used by fraudsters as a tactic in all of the other examples listed; a romance scam could be executed by the promise of a false investment opportunity (for example a 'pig butchering' scam[1298]) rather than a request for an assistive payment; a stolen identity might be used to carry out other types of fraud via user-generated content through a hacked account.

11.13 Fraud is the most frequently experienced crime in the UK.[1299] Fraud is also generally underreported, with estimates suggesting that fewer than one in seven instances (less than 14%) of fraud were reported to the police or Action Fraud (the public-facing national fraud and cybercrime reporting centre).[1300] Fraud victims have different reasons for not reporting, such as thinking that it would not help reduce the frequency of this crime, not knowing how to report it, feeling their experience was not worth reporting, or feeling ashamed or embarrassed.[1301] Ofcom research found that nearly nine in ten adult internet users (87%) have encountered content online that they believed to be a scam or fraud.[1302]

11.14 Most of the evidence found for relevant fraud offences relates to fraud by false representation, and that is therefore our primary area of focus in this chapter.[1303] However, other fraud offences can manifest online, including making or supplying articles for use in fraud. For more details on the fraud offences and how services can assess whether content amounts to illegal content, refer to the ICJG.

11.15 Ofcom research found that of the 11 types of scams or fraud tested during quantitative research, *"impersonation fraud (51%) was the most common type that had been experienced, followed by counterfeit goods scams (42%), investment, pension or 'get rich quick' scams (40%) and computer software service fraud or ransomware scams (37%)*".[1304]

11.16 Ofcom's Online Experiences Tracker found that most of the harms eliciting the highest level of expressed concern were encountered at a relatively low claimed incidence. However,

---

[1297] Research carried out by Which? reflects that user profiles, 'pages' and user groups on social media services are being used by criminals to provide stolen credentials, enabling the perpetration of further frauds through identity theft. "*They advertised a mixture of stolen identities, credit card details, compromised Netflix and Uber Eats accounts, as well as fraud 'how to' guides and fake passports made to order.*" Source: Which? (Lipson, F.), 2020. Your life for sale: stolen bank details and fake passports advertised on social media. [accessed 1 August 2023].

[1298] BBC One Rip Off Britain, 2023. What is a 'pig-butchering' scam – and why is it on the rise?. [accessed 18 October 2024].

[1299] National Crime Agency (NCA), 2024. Fraud. [accessed 9 September 2024].

[1300] Office of National Statistics (ONS), 2023. Crime in England and Wales: year ending June 2023. [accessed 31 July 2024].

[1301] UK Parliament The Parliamentary Office of Science and Technology (POST) (Low, N. and Lally, C.), 2024. Social and psychological implications of fraud. [accessed 31 July 2024].

[1302] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1303] Section 2 of the Fraud Act 2006 [accessed 2 August 2023].

[1304] The research asked participants about the different types of online fraud or scams they experienced if they self-reported to have personally been drawn into engaging with fraud or scams that began online. Participants may not have lost money but they may have, for example, clicked on an advertisement, followed specific instructions, or replied to a message. Source: Ofcom, 2023, Online Scams & Fraud Research. [accessed 2 August 2023].

'*scams, fraud and phishing*' is an exception, with relatively high levels of concern (79%) and experience (33%).[1305]

11.17 A large and increasing proportion of fraud involves the use of online services. Action Fraud, UK's national reporting centre for fraud and cybercrime, recorded £2.35bn as lost to fraud in 2021 to 22. It was identified that 80% (4 out of 5 instances) of reported fraud is cyber-enabled and that *"social media and encrypted messaging services as an enabler is increasing throughout all aspects of fraud"*.[1306] Money mules are also noted as a persistent feature across most fraud types (see also the 'Proceeds of crime' chapter). We are not able to determine the extent of the overlap between the 'social media and encrypted messaging services' and 'cyber-enabled' categories in the scope of the Act.

11.18 While it is challenging to estimate the economic and social cost of fraud, the 2023 UK Government Fraud Strategy estimates that the total economic and social cost of fraud to individuals between 2019 and 2020 to be £6.8bn[1307], significantly greater than simply the money lost to fraud.[1308] The UK government estimated the economic and social cost per fraud incident to be £1,427. Fraud types can overlap and use similar or the same tools to trick victims and gain their trust.[1309] Many who have experienced fraud say that it is common for many types of fraud to be contained in a single scam. For instance, a romance fraud scam may also involve elements of impersonation and investment fraud.

11.19 Ofcom research found that of those who had experienced scams or fraud, more than two in five (41%) said their last experience involved more than one type of scam or fraud.[1310] For example, impersonation fraud is one of the most sophisticated fraud types. In one scenario, the fraudster pretends to be a well-known brand on social media and uses a 'false' URL link that has a website domain closely related to the real domain that may be hard for consumers to distinguish, leading potential victims to a fake website. The fake website could look very similar, or almost identical, to the real website, persuading consumers to make a purchase that would not be fulfilled. Buyers' confidential payment details are likely harvested for further malicious purposes.[1311] Research from the DNS Research Federation found that when consumers are shown a URL with the presence of a brand name anywhere in the domain or subdomain, users would be more likely to trust and believe that it is legitimate.[1312]

[1305] Ofcom, 2024. Online Experiences Tracker - Wave 6. [accessed 18 October 2024].
[1306] Action Fraud, n.d. Fraud Crime Trends. [accessed 5 September 2023].
[1307] Note that the overall cost to the UK is considered to be significantly greater than just the value of the money reported lost directly to frauds. Source: Home Office, 2023. Fraud Strategy. [accessed 10 October 2024].
[1308] In comparison, UK Finance estimated to total money lost to fraud was £1.17 billion in 2023. Source: UK Finance, 2024. Annual Fraud Report 2024. [accessed 10 October 2024].
[1309] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].
[1310] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].
[1311] The Cyber Resilience Centre (Duckett, S.), 2023. What is Brand Impersonation? How can I Protect my Business?. [accessed 31 July 2024].
[1312] DNS Research Federation (Taylor, E. and Taylor, L.), 2023. Anatomy of a scam. What makes consumers click?. [accessed 31 July 2024].

Financial services offences: Examples

11.20 Examples of financial services offences include:

- A company or individual posting an unauthorised financial promotion; and
- A company or individual making a false claim to be authorised as a broker, for example, to potentially defraud victims.

11.21 Breaches of these rules may be committed by an individual or a provider intending to provide goods or services, or by fraudsters who are seeking to defraud victims. In the latter case, this would also constitute fraud by false representation and could be considered an investment scam or purchase scam as per the provided examples.

11.22 Consumers have increasingly been exposed to risk via unlawful financial promotions on services such as social media services.[1313] The Financial Conduct Authority (FCA) issued 1,882 alerts[1314] relating to unauthorised activity on its Warning List in 2022, up by 34% from 1,410 in 2021.[1315]

11.23 For more details on the financial services offences and how services can assess whether content amounts to illegal content, refer to the ICJG.

Risks of harm presented by fraud and financial services offences

11.24 The risks of harm to individuals from fraud is broad. Financial loss to victims is not the only result of fraud; the harm can be multi-faceted and can affect both mental and physical health. Recent research by Ofcom found that a quarter (25%) of those who had encountered an online scam or fraud lost money as a result, while more than a third (34%) reported that the experience had had an immediate negative effect on their mental health.[1316]

11.25 A report from Action Fraud said that reported losses due to fraud in 2020 to 2021 amounted to £2.35bn. It also found that adults aged 20 to 29 experience the highest frequency of fraud, but the greatest fraud-related losses are experienced by 50 to 69-year-olds.[1317]

11.26 A qualitative study by Ofcom showed that effects on scam victims continue after the scam ends. The research found that victims who had experienced scams encountered many challenges in their everyday lives. For example, anxiety and shame prevented them going to work or functioning in society; they were often cautious and wary when interacting with content and people online. Some lose confidence in their decision-making, feeling disappointed in themselves and become less trusting of other individuals.[1318]

[1313] DRCF, 2023. 2023/24 Workplan. [accessed 1 August 2023].
[1314] An alert warns people of unauthorised firms and individuals who are conducting unregulated activity. A list of these firms and individuals can be viewed on the FCA's warning list. Financial Conduct Authority (FCA), 2023. FCA Warning List of unauthorised firms. [accessed 1 August 2023].
[1315] FCA, 2022. Financial promotions data 2022. [accessed 1 August 2023].
[1316] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].
[1317] Action Fraud, 2021. Fraud Crime Trends. [accessed 1 August 2023].
[1318] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

11.27 Evidence presented by Age UK discusses how older scam victims, as well as people close to those who have been scammed, can experience digital exclusion due to their subsequent concerns about using the internet. This reduces their access to services, and among other disadvantages, can mean they pay more for certain goods.[1319]

11.28 The risks of harm to individuals from financial service offences is varied. People who are recruited to be money mules are very often unaware of the consequences and ultimately become victims. Mules who have been recruited sometimes suffer severe effects such as loss of access to financial services[1320], losing their homes and livelihoods and, in some cases, being a mule has even led to suicide.[1321]

Evidence of risk factors on user-to-user services

11.29 We consider that the risk factors below are likely to increase the risks of harm relating to fraud and financial services offences.

Risk factors: Service types

11.30 Research indicates that fraudsters use many types of U2U services. These include social media services, messaging services, marketplaces and listings services, and dating services. Most of the evidence for the offences relate to relevant types of fraud by misrepresentation, and we will identify where this relates to other fraud offences or the financial services offences.

Social media services

11.31 Our evidence reveals that social media services can be used by fraudulent actors in various ways. Ofcom research found that 15% of users' (15 in every 100) most recent experiences of impersonation fraud online originated on social media services, and nearly 3 in 10 (28%) counterfeit goods scams online were first encountered on social media services.[1322] In 2021, Action Fraud found that during a 12-month period, 5,039 reports of investment fraud referred to a social media service.[1323] Social media services can also be used to enhance the perceived legitimacy of fraudsters,[1324] and alongside dating services, are the primary enablers of romance fraud where scammers 'love bomb'[1325] their victims[1326]

11.32 According to the National Fraud Intelligence Bureau or NFIB, social media services or 'social networking sites' and online dating services are the main enablers of romance scams.[1327]

[1319] Age UK, 2023. Age UK- written evidence (DCL0049). [accessed 1 August 2023].
[1320] *"This can include bank account closure, limited access to loans or credit cards, difficulty obtaining a phone contract, and/or a prison sentence of up to 14 years."* House of Lords, 2022. Fighting Fraud: Breaking the Chain. [accessed 1 August 2023].
[1321] VICE (via YouTube), 2022. The Rise of Money Launderers on Snapchat and Instagram, 25 October. [accessed 2 August 2023].
[1322] Ofcom, 2023. Online Scams & Fraud Research: Data Tables. [accessed 2 August 2023].
[1323] Action Fraud, 2021. New figures reveal victims lost over £63m to investment fraud scams on social media. [accessed 1 August 2023].
[1324] National Fraud Intelligence Bureau (NFIB), 2022. Annual Assessment.
[1325] *"A romantic partner showers you with attention, money, and gifts in order to gain control in a relationship."* Reader's Digest (Nelson, B.), 2022. Is Love Bombing the Newest Scam to Avoid?. [accessed 1 August 2023].
[1326] The Guardian (Clark, J. and Wood, Z.), 2023. Victims speak out over 'tsunami' of fraud on Instagram, Facebook and Whatsapp. [accessed 1 August 2023].
[1327] NFIB, 2022. Annual Assessment.

Romance scammers seek to make direct contact with their victims and may look to move conversations to a private messaging service with encryption. They may 'love bomb' victims with frequent messaging and wait for many months before executing the scam.

11.33 Instances of 'cloned company investment fraud'[1328] and the use of social media in investment fraud[1329] has increased. In response to our 2022 call for evidence, the City of London Police said that Cloned Company Investment Fraud (CCIF) occurs when suspects pose as a legitimate firm and exploit their name and brand with a view to persuading victims to transfer funds for what they believe to be a genuine opportunity.[1330]

11.34 Action Fraud's Annual Assessment of Fraud Crime Trends in both 2020 and 2021 identified social media services and private messaging services with encryption as the primary enabler of all frauds, and advertising via search engine optimisation[1331] as another source of threat resulting in fraud and scams.[1332]

11.35 Research by Advocating against Romance Scammers found that fraudsters would join or follow groups on social media services which shared tips about various types of fraud, including those facilitated by identity theft[1333] and romance fraud.[1334]

11.36 An investigation by Which?, identified several profiles, pages and groups across multiple social media services by *"searching just a few slang terms used by fraudsters"*. These profiles, pages and groups *"advertised a mixture of stolen identities, credit card details, compromised Netflix and Uber Eats accounts, as well as fraud 'how-to' guides and even fake passports made to order"*.[1335]

Messaging services

11.37 Fraudsters use messaging services, including those with encryption to avoid detection and moderation.[1336] They often move conversations, started on other services, to messaging services with encryption.[1337] The increased level of privacy and few verification methods make encrypted messaging services attractive to fraudsters.

11.38 Research shows that functionalities that are central to many messaging services, such as direct messaging and group messaging, are risk factors for fraud.[1338] Ofcom research found

---

[1328] FCA, 2021. FCA issues warning over 'clone firm' investment scams. [accessed 2 October 2023].
[1329] Investment fraud can relate to financial services offences.
[1330] City of London Police response to 2022 Call for Evidence: First phase of online safety regulation.
[1331] The process of improving your site to increase its visibility when people search for products or services related to your business in Google, Bing, and other search engines. Search Engine Land, n.d.. What Is SEO – Search Engine Optimization?. [accessed 1 August 2023].
[1332] City of London Police response to 2022 Call for Evidence: First phase of online safety regulation. *"Social media services are used in multiple fraud offences, there were 138,375 Action Fraud reports that featured a social media or communication service in 2021/22. A total of £555m of financial losses related to these reports. In 2020/21 the figure was 75,769 - this increase of 83% indicates the accelerating trend of offenders using social media services to target UK victims for fraud."*; Publication of consultation on standalone code on fraudulent advertising to follow.
[1333] Identity theft relates to fraud by false representation.
[1334] Advocating Against Romance Scammers (Denny, B. and Waters, K.), 2021. Community Substandards: Capturing the Empty Promises of Big Tech's Safety against Online Romance Scams. [accessed 1 August 2023].
[1335] Which?, 2020. Your life for sale: stolen bank details and fake passports advertised on social media. [accessed 13 September 2023].
[1336] NFIB, 2022. Annual Assessment.
[1337] An NFIB profile into romance fraud in 2019 found that of the reports analysed, in 47% the conversation had been moved to a secondary encrypted service after the initial contract on a public primary service. City of London Police, 2019. 2019 Romance Fraud Profile, document owned by NFIB.
[1338] This may be relevant to offences for fraud and also for financial services.

that, according to survey respondents, just under half (46%) of fraudsters used a targeted message to make initial contact with their victim, and typically this is done through direct messaging (41%).[1339] The NFIB also found that fraudsters may also use group messaging to store information and to communicate with victims in a group.[1340]

Marketplaces and listings services

11.39 Ofcom research found that counterfeit goods scams, defined as those found at 'auctions and web marketplaces', were the second most-experienced type of scam among survey respondents.[1341] Meanwhile, data from UK Finance (a banking and finance trade body) indicated that purchase scams are the most common type of 'authorised push payment' fraud, and that these "*usually involve the victim using an online service such as an auction website or social media*".[1342]

Dating services

11.40 Our evidence points to fraudsters initially using dating services to find and target potential victims. Once a victim has been hooked, communication is likely to take place away from the original dating service, often being moved to an encrypted messaging service.[1343]

Risk factors: User base

User base size

Services with a large user base

11.41 Online services with a large user base are particularly attractive[1344] to fraudsters as they make it easy for them to reach large numbers of people inexpensively and with minimal effort. [1345]

11.42 In addition, a large user base makes it more likely that the initial reach of fraudulent UGC posts will be amplified to a bigger potential audience via a higher volume of content reactions, posts and re-posts.

11.43 Fraudsters make use of large, open user groups to add authenticity and to look for potential targets.

11.44 Having many user connections helps to add legitimacy to fraudsters and their content. Fraudsters have also used 'influencers' with large number of user connections to support their fraudulent work.

Services with a small user base

11.45 While larger services are a particular target for fraudsters, services with small user bases may also be targeted with some types of fraud. The NFIB has found that some fraudsters

---

[1339] Ofcom, 2023. Online Scams & Fraud Research: Data Tables. [accessed 2 August 2023].
[1340] NFIB, 2022. Annual Assessment.
[1341] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].
[1342] UK Finance, 2022. Annual Fraud Report. [accessed 1 August 2023].
[1343] Kaspersky, n.d. Online dating scams and how to avoid them. [accessed 1 August 2023]. Note that the source is a company specialising in cybersecurity.
[1344] Consumers International, 2019. Social Media Scams: Understanding the Consumer Experience to Create a Safer Digital World. [accessed 1 August 2023].
[1345] Federal Trade Commission (Fletcher, E.), 2022. Social media a gold mine for scammers in 2021. [accessed 1 August 2023].

look for more niche services in the UK, identifying services that are widely used by communities or professions which they can target.[1346]

11.46 For instance, romance fraudsters will join user groups centred around dating or making friendships such as widower groups or singles groups, and comment on their availability, compliment others, and seek to communicate privately. Fraudsters will also target investment groups; they often send mass messages, a practice which is less likely to be adopted by legitimate users looking for a personal connection.

User base demographics

11.47 The following section outlines crucial evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

11.48 Everyone is susceptible to fraud. Transient factors such as significant distraction, acute stress or serious emotional strain (beyond their emotional 'window of tolerance') can leave any person vulnerable which, in turn, can impair their ability to make sound decisions.[1347]

11.49 Data suggests that various personal characteristics play a role when it comes to identifying those who are more likely to experience risks and harm from different types of fraud. Someone's **age**, their **financial resilience**, **mental health**, and **media literacy** are all relevant characteristics, and differences within each of these are associated with cases of fraud. We'll explore these specific characteristics, but in every case of fraud, the unique characteristics of any one individual, and their circumstances at the time, will contribute to the risk they face and harm they experience.

Age

11.50 Although users of all ages can be victims of fraud, although different types of fraud, and the related content, affect different age groups differently.

11.51 For example, investment scams offering returns over time may gain a greater number of victims who are older and have disposable income, potentially from their pensions. Over-65s with more than £10k in savings are 3.5 times more likely to fall victim to a scam.[1348]

11.52 At the same time, in response to our 2022 call for evidence, the City of London Police said that social media has been a catalyst in the fraud victim pool becoming younger, with young people now losing money at a higher rate than older people, through investment and online shopping frauds. Younger people have spent a large proportion of their lives communicating online, so although they may be considered more tech savvy than older people, they may also be more open and trusting when sharing personal information in this space.[1349]

Financial resilience

11.53 The Phoenix Group (a long-term savings and retirement business based in the UK) found that "three in ten 18 to 34-year-olds fell victim to scams in the past year, with scammers

[1346] NFIB, 2022. Annual Assessment.
[1347] Which?, 2023. The Psychology of Scams. [accessed 30 August 2024].
[1348] FCA, 2016. Over 55s at heightened risk of fraud. [accessed 1 August 2023].
[1349] City of London Police response to 2022 Call for Evidence: First phase of online safety regulation.

turning to social media to target younger generations".[1350] Younger individuals (those aged 18 to 54) are the most likely to have low financial resilience (29% compared with the UK average of 24%).[1351] This may also mean they are susceptible to the types of fraud that appeal most strongly to those who are financially disadvantaged; where the supposed gains are promised quickly; purchase scams offering cheap goods; and loan-fee fraud scams which appear to peak at periods of financial difficulty.

11.54 The NFIB suggested that fraudsters were creating investment fraud opportunities online tailored to low-capital individuals, creating fraudulent offers that were tempting and accessible to a range of users with differing economic backgrounds, but who all use the same online services.[1352]

11.55 The FCA's Financial Lives Survey findings suggest that individuals from minority ethnic groups are more likely to have lower financial resilience,[1353] which may mean they are more susceptible to certain fraud types.

Mental health

11.56 Mental health conditions can put individuals at increased risk of becoming a victim of an online scam. A report from Money and Mental Health showed that "*people who have experienced mental health problems are three times more likely than the rest of the population (23% versus 8%) to have been a victim of an online scam*". The report also says that those with "*impaired decision-making, increased impulsivity and low motivation can all make it difficult for people with mental health problems to spot fraud and avoid losing money or personal information*".[1354]

Media literacy

11.57 The media literacy level of an individual may influence whether they can recognise fraud. Ofcom's media literacy work shows that there is often a gap between people's high confidence that they can identify scam messages, and their actual (low) ability to do so.[1355]

11.58 For example, research considering internet users' ability to spot impersonators or copycat accounts shows that users have varying levels of competence in judging whether user profiles or 'accounts' are credible. Looking at account verification, 28% (less than one third)

[1350] Phoenix, 2021. Three in ten 18-34 year olds fell victim to scams in the last year, with scammers turning to social media to target younger generations. [accessed 20 September 2023].

[1351] "*Adults are described as having low financial resilience if they have little capacity to withstand financial shocks, because, for example, they do not think they would be able to withstand losing their main source of household income for even a week or are finding it to be a heavy burden keeping up with their domestic bills or credit commitments, or because they have already missed paying these bills in 3 or more of the last 6 months. So, our definition includes both those adults who are already in financial difficulty (because they are missing bills – so this is an objective measure) and those who could quickly find themselves in difficulty if they suffer a financial shock (by more subjective measures)*". FCA, 2022. Financial Lives 2022 survey: insights on vulnerability and financial resilience relevant to the rising cost of living. [accessed 1 August 2023].

[1352] NFIB, 2022. Annual Assessment.

[1353] This is compared with the UK average (Ethnicity: Black & Black British 44%, Mixed/Multiple 39%, UK average 24%). FCA, 2022. Financial Lives 2022 survey: 3. Low financial resilience. [accessed 1 August 2023].

[1354] The Money and Mental Health Policy Institute (Holkar, M. and Lees, C.), 2020. Caught in the Web. [accessed 1 August 2023].

[1355] Ofcom, 2022. Adults' Media Use and Attitudes report. [accessed 2 August 2023]. Note: The research relates to paid advertising content (out of scope of this assessment), where four in ten claimed they would not be able to tell if an advert was fake or not.

of respondents to Ofcom research stated that they always, mostly, or sometimes checked for verification symbols when deciding to follow or interact with an account.[1356]

Risk factors: Functionalities and recommender systems

User identification

User profiles

11.59 Services which allow individuals to create user profiles quickly and easily can be exploited by fraudsters to establish an online presence which looks and feels legitimate or credible to other users. Creating a user profile is also attractive to fraudsters as it does not require them to undertake additional steps such as designing a website or paying for website hosting.

11.60 User profiles and the information that they often display can be also used as a tool by fraudsters to gather information about potential victims. The NFIB found that, depending on the level of restriction, user profiles can provide fraudsters with information about individuals which could be exploited or cross-referenced with data elsewhere. For instance, job and career histories on user profiles can help fraudsters to identify high net-worth individuals.[1357]

11.61 User profiles have also been misused by criminals seeking to attract other criminals or *"would-be ID fraudsters"* into communicating with the intention of obtaining or offering to supply articles for use in frauds. A Which? investigation into the sale of stolen bank details found *"50 scam profiles, pages and groups"* on various social media services.[1358]

11.62 Compromised social media accounts can be utilised by fraudsters. This includes regular user accounts, to carry out purchase scams for example[1359]. However, accounts labelled with a 'verified' status are more likely to facilitate a scam compared to unverified accounts, because such an official mark of verification enhances perceived credibility.[1360] Verified stolen accounts have been found to be on sale online[1361] and subsequently used to post content via the verified profile to defraud users.[1362] There are examples of both individuals' and institutions' verified accounts being hacked, with subsequent account name and profile changes made to perpetrate a scam.[1363]

[1356] Ofcom, 2023. Open Data. [accessed 5 September 2023].
[1357] NFIB, 2022. Annual Assessment.
[1358] Which? (Lipson, F.), 2020. Your life for sale: stolen bank details and fake passports advertised on social media. [accessed 13 September 2023].
[1359] Youngs, I., 2024. Facebook 'did nothing about Taylor Swift ticket hack scam', BBC News, 8 May. [accessed 18 October 2024].
[1360] Morris, M. R., Counts, S., Roseway, A., Hoff, A. and Schwarz, J., 2012. Tweeting is believing? Understanding microblog credibility perceptions, *Computer Supported Cooperative Work: Proceedings of the ACM 2012 Conference*. [accessed 31 July 2024].
[1361] Agarwal, S., 2022. The black market for stolen verified accounts from Twitter and Instagram, The Verge, 18 October. [accessed 31 July 2024].
[1362] Francisco, E., 2020. July 15 Twitter hack: A list of every hacked verified account, Inverse, 20 February. [accessed 31 July 2024].
[1363] Binder, M., 2023. Scammers hack verified Facebook pages to impersonate Meta and Google, Mashable, 5 May. [accessed 1 August 2024].

Fake user profiles

11.63 Fake user profiles can be used by fraudulent actors to conceal their identity and impersonate notable entities such as banks, insurance providers or high-profile people in the finance sector. For instance, 14 public figures co-signed a letter in 2021, discussing how their images had been used by fraudsters to exploit victims' trust.[1364]

11.64 The fraud prevention service Cifas stated that a primary contributing factor to the scale and resulting harm of online fraud and wider crimes, is the lack of effective verification of user accounts on social media services. Cifas said: "*this enables criminals to hide behind the anonymity of fake profiles when targeting victims, and to impersonate trusted sources".*[1365]

11.65 Clean Up the Internet[1366] has reported on "*how fraudsters exploit fake social media accounts to scam UK users*".[1367] Ofcom qualitative research with online users who said they had encountered online fraud highlighted that "*social media has become a common way for companies and brands to communicate with potential customers, and scammers are taking advantage of that to make contact with potential victims*".[1368]

11.66 Romance fraudsters use the guise of genuine relationships to manipulate victims for financial gain and other potential criminal activity. They do this by creating fake personas on dating sites and social media services.[1369] In its response to our 2022 call for evidence, the City of London Police said "social media has become another method of contacting victims using profiles with stolen images. Offenders identify potential victims and go on to send connection requests and messages, and quickly encourage conversations via less regulated and encrypted messaging services".[1370] The impact on victims can be substantial. A case study found that one fraudster had conned 80 victims out of over £400,000 between 2005 and 2021. One of these victims was defrauded over a period of 14 years, during which they gave the fraudster over £100,000.[1371]

11.67 Evidence has shown that if users can get their account 'verified' (for example obtaining a lookalike label as a verified account status by paying a fee), such an approach is likely to be exploited by scammers.[1372] The feasibility of paying for a badge that looks the same as a label that has been established as trustworthy in the past, means a vast range of figures and brands can be impersonated easily, where additional processes or checks are not in place.[1373]

[1364] King, S., 2021. Martin Lewis, Sir Richard Branson, Deborah Meaden and other public figures issue plea to the PM to put scam ads in the Online Safety Bill, MoneySavingExpert News, 16 November. [accessed 1 August 2023].
[1365] Cifas response to 2022 Call for Evidence: First phase of online safety regulation.
[1366] Clean Up the Internet is an independent, UK-based organisation concerned about the degradation in online discourse and its implications for society and democracy.
[1367] Babbs, D, 2023. New report on fraud, fake accounts, and the User Verification Duty, *Clean Up the Internet*, 25 April. [accessed 1 August 2023].
[1368] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].
[1369] LoveSaid response to November 2023 Illegal Harms Consultation, LoveSaid (Ofcom.org.uk); College of Policing (Cumming, L.), 2021. Romance fraud: Five things you need to know. [accessed 31 July 2024].
[1370] City of London Police response to 2022 Call for Evidence: First phase of online safety regulation.
[1371] Home Office, 2023. Fraud Strategy: Stopping Scams and Protecting the Public. [accessed 8 August 2024].
[1372] Burgess, M., 2022. Elon Musk's Twitter Is a Scammer's Paradise, Wired, 10 November. [accessed 1 August 2024].
[1373] Clean up the Internet (Kinsella, S.), 2022. What do Elon Musk's "Blue Tick" experiments mean for the UK's Online Safety Bill?. [accessed 1 August 2024].

User networking

User connections

11.68 The ability to accrue user connections with little friction creates a risk of services being used to commit or facilitate fraud offences as it allows fraudsters to build up large followings. This can add authenticity to their online presence. Consequently, user connections can help enhance the perceived legitimacy of fraudsters and their content.[1374]

11.69 Fraudsters have also used notable entities or 'influencers'[1375] with significant number of followers to help to facilitate fraud. The NFIB found that this could be done in a number of ways; for instance, by using the image or intellectual property of influencers in 'edited advertisements'[1376] to falsely imply an association and add legitimacy to their activities; by setting up fake user profiles claiming to be affiliated with the influencers; and in some cases by convincing the influencer to advertise on their behalf, believing that the opportunities are credible.[1377]

11.70 The FCA have observed harm from illegal financial promotions posted by influencers on social media platforms. This can include "*archetypal celebrity influencers who are not associated with financial services but have large follower groups*".[1378]

User groups

11.71 Research shows that fraudsters join or follow groups on social media services where tips about various types of fraud are shared, to learn about targets and potential scams.[1379] User groups also make it easier for fraudsters to identify and connect with potential targets.[1380]

11.72 Research by Which? also found examples of user groups "*promoting identity theft and other types of fraud*" on social media sites. These groups hosted content advertising stolen identities and credit card details.[1381]

11.73 The FCA also note the role of influencers who may be communicating illegal financial promotions in "forums and discussion groups on financial topics that function as spaces in which individuals exchange information and share knowledge. These groups are set up to encourage participants to register for a specific course or are used by participants to

---

[1374] NFIB, 2022. Annual Assessment; A study into fake animal rescues noted that a large following is potentially lucrative for fraudsters who seek to monetise on 'fake' rescue content or build legitimacy for donations or other rewards. Social Media Animal Cruelty Coalition, 2024. Spot the Scam: Unmasking Fake Animal Rescues. [accessed 29 October 2024].
[1375] Influencers gather large followings of enthusiastic, engaged people who pay close attention to their views.
[1376] NFIB, 2021. The Role of Social Media in Investment Fraud. Examples of 'edited advertisements' include but not limited to content where images of notable entities or 'influencers' were extracted from existing media and added to fraudulent advertisements made by fraudsters: Sproson, K. and Slater, B., 2024. Martin Lewis scam adverts, MoneySavingExpert, 7 October. [accessed 18 October 2024]; King, S., 2021. Martin Lewis, Sir Richard Branson, Deborah Meaden and other public figures issue plea to the PM to put scam ads in the Online Safety Bill, MoneySavingExpert News, 16 November. [accessed 1 August 2023].
[1377] NFIB, 2022. Annual Assessment.
[1378] FCA, 2024. FG24/1: Finalised guidance on financial promotions on social media. [accessed 10 September 2024].
[1379] Advocating Against Romance Scammers (Denny, B. and Waters, K.), 2021. Community Substandards: Capturing the Empty Promises of Big Tech's Safety against Online Romance Scams. [accessed 1 August 2023].
[1380] For example, fraudsters used social media groups to identify targets trying to buy Taylor Swift tickets, who they could then try and scam. BBC News (Smith, E. and Horsburgh, L.), 2024. Taylor Swift ticket scammers 'feed off fans' desperation'. [accessed 24 September 2024].
[1381] Which? (Lipson, F.), 2020. Your life for sale: stolen bank details and fake passports advertised on social media. [accessed 13 September 2023].

encourage others to engage in personal chats outside the platform where they sell financial advice or financial products".[1382]

11.74 UK Finance note the false sense of community that criminals can create by posting false claims of success from investments which, in turn, can encourage individuals to fall victim to scams.[1383] Criminals can also act as moderators of user groups to enhance their credibility and perpetrate scams, as noted by a BBC investigation into sales of fake Taylor Swift concert tickets.[1384]

User communications

Livestreaming

11.75 Financially motivated sexual extortion, also referred to as 'sextortion', can be carried out over livestreams.[1385] Sometimes perpetrators use 'fake identities' in these livestreams, presumably by creating inauthentic user profiles, which can be regarded as **fraud by misrepresentation**.[1386] The NCA says that "*criminals might befriend victims online by using a fake identity and then then trick them into performing sexual acts in front of their webcam. These webcam videos are recorded by the criminals who then threaten to share the images with the victims' friends and family*."[1387] The NCA adds that both men and women can be victims of this crime, either by being blackmailed or by being coerced into carrying out sexual acts.

Direct messaging

11.76 Direct messaging can be an enabler of fraud. Ofcom research found that, according to survey respondents, just under half (46%) of fraudsters use a targeted message to make initial contact with their victim, and typically, 41% of this is done through direct messaging.[1388] Depending on their selected settings, users can receive messages from others they may not know, which can lead to a risk of harm.

11.77 This same research also found that fraudsters often employed several techniques to gain victims' trust. The report says*: "the scammer often employed one or more engagement techniques which impairs the rational decision-making process such as:* constant contact and messaging victims*, telling hardship tales, giving victims a return on their initial investment, being charming, or emphasising time sensitivity (for example, 'to get this price you need to sign up in the next 24hrs')*.[1389]

[1382] FCA, 2024. FG24/1: Finalised guidance on financial promotions on social media. [accessed 10 September 2024].
[1383] UK Finance response to November 2023 Illegal Harms Consultation.
[1384] BBC Radio 4 You and Yours (Vahl, S. and Smith, E.), 2024. Facebook Ticket Scam, Business Nimbys and Smart Meter Update. [accessed 24 October 2024].
[1385] 'Financially Motivated Sexual Extortion' is a form of blackmail that involves threatening to publish sexual information, photos or videos about someone.
[1386] Please note that where the victim of financially motivated sexual extortion is a child then the activity and content would also likely constitute one or more grooming or CSAM offences. For details, see 'Financially motivated sexual extortion' in the CSEA chapter.
[1387] National Crime Agency, n.d. Kidnap and Extortion. [accessed 1 August 2023].
[1388] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].
[1389] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

11.78 Fraudsters also use the direct messaging function on social media services to engage with other fraudsters when seeking to obtain articles for use in fraud.[1390] As part of an investigation for BBC Panorama, an investigative journalist engaged directly with a fraudster and obtained access to information on how to commit fraud and where to buy personal and financial details.

Group messaging

11.79 Group messaging may also be an enabler, allowing fraudsters to communicate with many potential victims at once. The NFIB found that fraudsters may use such messaging facilities to store information and to communicate with victims in a group, creating an appearance of an organised business practice while allowing the fraudster to target multiple victims with ease.[1391]

Private messaging and encrypted messaging

11.80 Private and encrypted messaging services are inherently attractive environments for fraudsters, both as a location to commit or discuss fraud, as well as a destination to migrate potential victims who have been initially approached in other online spaces.[1392] Fraudsters will use private messaging services with encryption to avoid anyone (including the service itself) moderating their conversations or searching for 'red flags' which may disrupt their activity. The evidence gives insight into this technique used for fraud offences.

11.81 In response to our 2022 call for evidence, the City of London Police said that "A popular tactic for dating scammers is to move the conversation from dating services with an increased level of support to messaging services with end-to-end encryption. These encrypted services become even more attractive for fraudsters with new security features being added, which hinder police investigation, such as disappearing messages and notifications when individuals screen shot conversations".[1393]

11.82 In response to our 2022 call for evidence, the City of London Police also said that "Encrypted services are attractive to fraudsters due to an increased level of privacy. As well as individual messaging, scammers can use services to target and create groups of intended victims under the guise of a legitimate business practice. These accounts are easy to set up with less verification needed than on other services".[1394]

Commenting on content

11.83 The ability to comment on content may enable fraud, although in some contexts these can also help to flag fraudulent activity to other users.

11.84 Ofcom research found that one in 20 (5%) UK internet users who had encountered scams, fraud or phishing in the past four weeks had first encountered it in the comments or replies to a post, article or video.[1395] As well as being a source of fraudulent content, it appears that comments also play a role in helping users to recognise fraudulent content. Eighteen

---

[1390] Okpattah, K., 2021. Social media fraud: The influencers promoting criminal scams, BBC News, 16 August. [accessed 13 September 2023].
[1391] NFIB, 2022. Annual Assessment.
[1392] NFIB, 2022. Annual Assessment.
[1393] City of London Police response to 2022 Call for Evidence: First phase of online safety regulation.
[1394] City of London Police response to 2022 Call for Evidence: First phase of online safety regulation.
[1395] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].

percent (just under 1 in 5) of those who had encountered scams or frauds said they were suspicious because of comments from other users voicing concerns.[1396]

11.85 Research by Cifas and Forensic Pathways found that criminals advertise the sale of personal details on forums due to the *"enhanced level of exposure",* noting that "*forums on the surface web are more easily accessible than those on the dark web and therefore the possibility of more people seeing such posts is heightened. There is also a high turn-over of messages posted on forums*". The research suggests that the ability to post comments or messages on an open forum increases accessibility to target audiences, for the purpose of supplying or offering to supply articles for use in frauds, for example guidebooks or stolen credentials.[1397]

Hyperlinking

11.86 Hyperlinks play an important role in facilitating purchase scams and advance fee scams. [1398] Hyperlinks can also be embedded in communications to victims who are encouraged to click the hyperlink which can redirect them to unexpected webpages, which, in turn, can be used to trick visitors into divulging personal information or cause malicious programmes to be downloaded onto someone's device.[1399] This approach is often associated with the harvesting of credit or debit card credentials and user information. This information can be used for social engineering of victims at a later date.[1400]

11.87 Which? also note that suspicious URLs can be a good indicator of the existence of fraudulent content, although, alone, do not provide enough grounds to infer the content being hyperlinked to is fraudulent.[1401]

Transactions and offers

Posting goods or services for sale

11.88 The ability to post goods and services online has been exploited by fraudsters, and the evidence is associated to fraud offences.

11.89 Data from UK Finance indicates that purchase scams are the most common type of 'authorised push payment' fraud, and that these *"usually involve the victim using an online service such as an auction website or social media*".[1402] While some services which offer the opportunity to post products for sale have secure payment options, fraudsters may prefer direct bank transfers.

11.90 Being able to post goods or services for sale is an important feature for committing purchase scams in which victims pay for goods or services which never arrive or are counterfeit.[1403] Common examples include criminals posing as sellers of products like

[1396] Ofcom, 2023. Online Scams & Fraud Research 2022. [accessed 2 August 2023].
[1397] Cifas and Forensic Pathways, 2018. Wolves of the Internet. [Accessed: 13 September 2023].
[1398] UK Finance response to November 2023 Illegal Harms Consultation. UK Finance (ofcom.org.uk).
[1399] McAfee (Dhaliwal, J), 2023. What Are the Risks of Clicking on Malicious Links. [accessed 10 October 2024]
[1400] UK Finance response to November 2023 Illegal Harms Consultation. UK Finance (ofcom.org.uk).
[1401] Which? response to November 2023 Illegal Harms Consultation. Which? (pfcom.org.uk).
[1402] UK Finance, 2022. Annual Fraud Report. [accessed 1 August 2023].
[1403] UK Finance, 2022.

computers and smartphones, or advertising for sale fake holiday rentals, concert tickets[1404], or exam papers. [1405]

11.91 A PDSA animal wellbeing report has also shown a growth in people buying their pet online, from 53% in 2022 to 65% in 2023 (equating to 15 million pets) which has exacerbated pet purchase scams.[1406] Even when the pet is real, scammers could use false information about the animal to mislead and defraud buyers.[1407] This could have repercussions on other criminal activity such as increasing the demand for smuggling animals from abroad.[1408] Victims reported being scammed on social media (11%), online marketplaces (25%) and specific pet-selling services (37%).[1409]

Content exploring

User-generated content searching

11.92 The ability to search for UGC can be an enabler of fraud. Fraudsters can use this functionality to search for potential victims and it can also act as a warning sign to others that the content may be fraudulent.

11.93 Ofcom research found that 4% of UK internet users who had encountered scams, fraud or phishing most recently in the past four weeks had first encountered it when watching content they had chosen to watch.[1410] Three percent said they had first encountered it when using the search function.[1411]

11.94 Other evidence suggests that fraudsters can find posts offering to supply articles and information which support the commission of fraud. A defining characteristic of this type of content is the dense combining of terms.[1412] A BBC report similarly identified 'fraud' influencers who openly post fraud articles such as stolen bank details alongside advice on how to use them to commit fraud. These posts can be easily found by other users through content searching.[1413]

11.95 Research completed by Ofcom[1414] as well as research by other organisations[1415] has shown that some social media services and search services are being used by criminals to supply articles for use in frauds with virtually no effort on the part of criminals to conceal their intentions.[1416] Further desk research suggests that, while it is unlikely to be encountered

---

[1404] BBC News (Smith, E. and Horsburgh, L.), 2024. Taylor Swift ticket scammers 'feed off fans' desperation'. [accessed 24 September 2024].
[1405] Johnson, K., 2024. GCSE pupils targeted by 'manipulative' exam scams, BBC News, 7 May. [accessed 24 October 2024].
[1406] PDSA, 2023. PDSA Animal Wellbeing (PAW) Report. [accessed 2 August 2024]; Lloyds Banking Group, 2023. Fraudsters go unleashed online as pet scams rise. [access 2 August 2024].
[1407] Cats Protection, 2023. CATS Report. [accessed 2 August 2024].
[1408] Dogs Trust, n.d. The Puppy Smuggling Scandal. [accessed 2 August 2024]; Sky News (Jones, T.), 2022. Dogs Trust warns people not to risk buying smuggled puppies this Christmas. [accessed 2 August 2024].
[1409] Action Fraud, 2021. Ruff time for animal lovers as scale of pandemic pet fraud unleashed. [accessed 2 August 2024]; BBC News (O'Donoghue, D. and Hesketh, S.), 2024. Missing pets: 'Heartless' scammers targeting desperate owners. [accessed 2 August 2024].
[1410] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].
[1411] Ofcom, 2024. Online Experiences Tracker – Wave 6. [accessed 22 November 2024].
[1412] Ofcom, 2023. Prevalence of Potentially Prohibited Items on Search Services. [accessed 21 September 2023].
[1413] Okpattah, K., 2021. Social media fraud: The influencers promoting criminal scams, BBC News, 16 August. [accessed 23 August 2023].
[1414] Ofcom, 2023. Prevalence of Potentially Prohibited Items on Search Services. [accessed 21 September 2023].
[1415] Okpattah, K., 2021.
[1416] SEON, n.d. What Are Fullz. [accessed 4 September 2023]; Fraud.net, n.d.. What is Fullz?. [accessed 4 September].

accidentally by internet users, this type of content is often very discoverable by criminals and likely to be prevalent on the open web and dark web – often on online forums.[1417] Once criminals have acquired access to a package of stolen financial credentials and related personal information, this access will then typically be used to undertake a wide range of secondary fraud activities. These include card-related fraud (for example, the fraudulent purchase of goods, services, or subscriptions, making payments to 'money mule' accounts to launder the proceeds of crime), or impersonation or identity fraud (stealing someone's identity to take over or set up new bank accounts, email accounts or social media profiles to support fraudulent loan applications.).[1418]

11.96 Similarly, research commissioned by Cifas in 2018 revealed that packages including personal data and financial information sell for about £31 on the surface web, while data held on the magnetic strip of bank cards sells for around £70.[1419]

Risk factors: Business models and commercial profiles

Revenue models

11.97 Ofcom research showed that counterfeit goods are often bought on online marketplaces and auction sites[1420] which involve charging a transaction fee as an essential part of their revenue model. Furthermore, online marketplaces and social media services often also offer a functionality where users can pay to promote their user-generated posts to a wider audience by the service. Transaction fees, resulting from marketplace purchases, and the 'pay to promote' functionality on services present a commercial incentive and generate revenue for services. However, marketplaces and the 'pay to promote' facility can also provide opportunities for fraudsters to create and promote fraudulent content.[1421]

[1417] Bodker, A., Connolly, P., Sing, O., Hutchins, B., Townsley, M. and Drew J., 2022. Card-not-present fraud: using crime scripts to inform crime prevention initiatives, *Security Journal*, 36 (693-711). [accessed 23 August 2023].
[1418] SEON, n.d.; Data Dome, 2023. What are fullz? How do fullz work?. [accessed 4 September 2023].
[1419] Cifas and Forensic Pathways, 2018. Wolves of the Internet. [accessed: 28 September 2023].
[1420] *"Counterfeit goods were described as fake designer brand clothes, accessories, perfumes, pirated copies of DVDs and computer games, often found at auctions and web marketplaces, where you can't check if the products are genuine until the item has been delivered".* Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].
[1421] We note that content that users pay to promote is within the scope of this risk assessment and the wider regime. There are separate duties for 'fraudulent advertising' that apply to non-user-generated content.

12. Proceeds of Crime

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for proceeds of crime priority offences: how harm manifests online, and risk factors

'Proceeds of crime' is the term used for money or assets gained by criminals during their criminal activity and money laundering. Examples of activities which involve the proceeds of crime online include people being recruited as money mules to transfer illegally obtained money between bank accounts, discussion between criminals to arrange money laundering, and stolen personal information (via other criminal activity) offered for sale which can be used to commit or facilitate other types of fraud.[1422] Further information on fraudulent activity can be found in the chapter 'Fraud and financial services offences'.

'Proceeds of crime offences' is taken to mean offences relating to the concealment, arrangement of, acquisition, possession and use of criminal property in the Proceeds of Crime Act 2002.

The risks of harm to individuals that could arise from proceeds of crime offences include losing livelihoods, losing access to financial services and consequential effects on mental health.

Service type risk factors:

Our evidence shows that proceeds of crime offences committed or facilitated online most commonly rely on **social media services**, **messaging services** and **gaming services** that have a broad userbase and wide reach.

User base risk factors:

Online services with **large user bases** are particularly attractive to fraudsters (including money mule recruiters) as they make it easy for them to reach large numbers of people at low cost, with minimal effort. In addition, a large user base can make it more likely that the initial reach of fraudulent content will be shared with a larger audience through likes, shares and re-shares.

Services which make use of **large, open groups of users** will also be attractive to mule recruiters as they tend to add legitimacy to criminals. Services with a large user base are more likely to provide such groups.

[1422] A money mule, or simply 'mule', '*is someone who receives money from a third party in their bank account and transfers it somewhere else, or who withdraws it as cash and gives it to someone else, obtaining a commission for it or payments in kind*.' These individuals are targeted by 'money mule recruiters', sometimes referred to as 'mule herders', who recruit money mules, often using social media or online gaming services. House of Lords, 2022. Fighting Fraud: Breaking the Chain. [accessed 25 September 2023].

Our evidence also shows that individuals from **lower-income households** and **from certain minority groups** may be more at risk of harm from this offence. Young adults were also seen to be more at risk as they sometimes have lower financial resilience and are more likely to have "clean" accounts. This could lead them to fall victims to potential money mule recruiters with the promise of money.

Functionalities and recommender systems risk factors:

The ability to create **fake user profiles** can make it harder to trace money mule recruiters and individuals posting fake job opportunities.

Recruiters of money mules will use user-to-user (U2U) services to contact potential victims easily and directly. **Direct messaging** can be used by recruiters to directly contact potential money mules, often using specific phrases to attract people. The functionality of **user-generated content (UGC) searching** can also enable victims to initiate contact. Potential victims can respond to a post offering the chance to make money after searching for relevant content or seeing a misleading job opportunity.

Criminals intending to commit proceeds of crime offences are likely to prefer **encrypted messaging** to communicate between themselves. The **ability to post content** and **comment on posts** can also enable perpetrators to find and contact at-risk individuals.

User profiles, and the information displayed on them, can be used by perpetrators to gather information surrounding a potential victim.

Introduction

12.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to the proceeds of crime offences listed under 'Relevant offences'; and
- The use of these services for the commission or facilitation of these offences (collectively, the 'risks of harm').

12.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical or psychological harm as part of our assessment of the risks of harm, and where possible, consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

Relevant Offences

12.3 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding proceeds of crime offences, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

12.4 The priority offences for proceeds of crime are the following:

a) Concealing etc criminal property[1423]
b) Arrangements related to criminal property[1424]
c) Acquisition, use and possession of criminal property[1425]

12.5 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and in relation to offences in Scotland, being involved art and part in the commission of these offences).

12.6 For more details on the offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance or ICJG.

How proceeds of crime offences manifest online

12.7 This section is an overview which looks at how proceeds of crime offences manifest online, and how individuals may be at risk of harm.

12.8 Based on the currently available evidence, money muling and stealing of personal data via other criminal means are the primary, and most prevalent, ways in which proceeds of crime offences manifest via user-generated content (UGC) and user interactions, in scope of the illegal content safety duties under the Act. The City of London Police have also stated that "the continued use of 'money mule networks' to receive, move and conceal the proceeds of fraud is an ongoing and persistent threat evidenced across many fraud types, and continues to facilitate the movement of fraudulent funds as well as access to victims across domestic and international jurisdictions".[1426]

12.9 Examples of proceeds of crime activities online could include recruiting people as money mules to transfer illegally obtained money between bank accounts, enabling discussion between criminals to arrange money laundering, and sharing stolen personal information. Money mules are people who knowingly or unknowingly help criminal organisations launder their illicit profits. They do this by using their bank accounts to receive and transfer fraudulent funds, thereby making them appear legal.[1427]

12.10 Criminal organisations often rely on money laundering to conceal the origin of illicit funds. The aim is to make funds appear legitimate, at the same time distancing criminal groups from prosecution. This action also makes it difficult for law enforcement to trace money trails.

12.11 A variety of scenarios can lead to proceeds of crime offences in the context of UGC. Here we provide a non-exhaustive list of examples:

- Users may be tricked into becoming a money mule without being aware that they have become involved in illegal activity. For example, money mule recruiters may create fake jobs that involve moving money between accounts, including asking the victim to use their own account to help move the money, or to hand over control of their account. Other tactics used to hook potential victims could include a romance scam, where the

[1423] Section 327 of Proceeds of Crime Act 2002.
[1424] Section 328 of Proceeds of Crime Act 2002.
[1425] Section 329 of Proceeds of Crime Act 2002.
[1426] Evidence provided by City of London Police to the House of Lords 'Fraud Act 2006 and Digital Fraud Committee'. Source: House of Lords, 2022. Fighting Fraud: Breaking the Chain. [accessed 22 September 2023].
[1427] Interpol, n.d. Money mules – what are the risks? [accessed 22 September 2023].

money mule recruiter exploits the victim's trust to ask them to transfer money or share their account details.[1428]

- In other circumstances, mules may be aware of or partially complicit in potentially engaging in illegal activity – actively responding to opportunities to make money. These money mules may make transfers or agree to surrender control of their accounts in return for earning a commission.
- Alternatively, potential perpetrators can communicate with each other through UGC. This may include criminals coordinating or arranging money laundering via private messaging or using UGC posts to promote the sale of stolen financial or personal credentials[1429] which can be used to launder funds.

12.12 Mule recruiters will use bank facilities such as bank transfers to transfer money from one online account to another. A paper from Sanction Scanner, a firm working in anti-money laundering, points out that "criminals choose an account that does not have a criminal record to reduce the likelihood of getting caught while choosing money mules. The money to be laundered is transferred from the mule account to the third-party bank account via bank transfer, and the money received is converted into cash. After that, this money is converted into a virtual currency like Bitcoin".[1430]

12.13 The National Crime Agency (NCA) identifies money muling as one of the most important enablers of fraud online. Fraudsters and other criminals will use mule accounts to make it harder for banks and law enforcement to track them down.[1431] Cifas, the Credit Industry Fraud Avoidance system, reported that in 2022 there were 39,578 cases of bank account activity indicative of money mule behaviour. Although UGC may not be a feature in all cases, social media has been identified as a "key enabler in the recruitment of mules".[1432]

12.14 Content and interactions can amount to, or facilitate, a proceeds of crime offence, regardless of whether the money mule is unaware or complicit in that activity.

12.15 In addition to money muling, the theft and dissemination of stolen credentials online is widespread, and while covered by other specific fraud offences, could be relevant to Proceeds of Crime offences. An investigation by the consumer advocacy body, Which?, found that it was 'rapidly' able to find stolen information on social media sites, including "identities, credit card details, compromised Netflix and Uber Eats accounts".[1433] The Crown Prosecution Service, the principal agency conducting criminal prosecutions in England and Wales, notes the existence of 'online marketplaces' used by criminals to sell stolen credit card details, among other items.[1434] Cifas has also flagged that identity fraud cases have reached "an all-time high as the cost-of-living crisis bites".[1435]

[1428] Triodos Bank, n.d. What is money muling? [accessed 22 September 2024]; UK Finance response to November 2023 Illegal Harms Consultation, Appendix 1.
[1429] Note that the sale of, or offering for sale, stolen personal details itself may constitute the offence of making or supplying articles for use in frauds. See the chapter 'Fraud and financial services offences' for further information.
[1430] Sanction Scanner, n.d. The Change of Money Laundering in The Digital Age. [accessed 22 September 2023].
[1431] FBI, n.d. Money Mules. [accessed 22 September 2023].
[1432] Cifas, 2023. Fraudscape 2023. [accessed 22 September 2023].
[1433] Which?, 2020. Your life for sale: stolen bank details and fake passports advertised on social media [accessed 22 September 2023].
[1434] Crown Prosecution Service, 2019. Cybercrime - prosecution guidance. [accessed 22 September 2023].
[1435] Cifas, 2023. Identity fraud cases reach all-time high as cost-of-living crisis bites. [accessed 22 September 2023].

12.16 Other crimes related to proceeds of crime offences identified by West Yorkshire Police include drug trafficking, human trafficking and cyber-crimes such as ransomware attacks.[1436] FBI Omaha have also reported cases of children and young adults being recruited to launder money that is used to buy links to child sexual abuse material (CSAM). In these cases, perpetrators pose as a fake company and ask individuals to accept money from the company's customers. Recruits can keep a fee and change the rest into cryptocurrency, which is then sent on to the fake company.[1437]

12.17 The Children's Society observe from their work that children "*are increasingly being groomed into illegal activities like money laundering*."[1438] Fake job adverts offer a promise of quick money, while they can also be asked to share bank details via social media and gaming platforms.

Risks of harm to individuals presented by online proceeds of crime offences

12.18 People who are recruited to be money mules are often unaware of the consequences and can ultimately become victims. A Crimewave video on the NCA website looked at the rise of money laundering on social media sites. It showed that money mules sometimes suffer severe effects: losing their homes and livelihoods, losing access to financial services[1439] and suffering diminished mental health.[1440]

12.19 Once a person becomes a money mule, it can be very difficult for them to stop. Cifas points out that money mules "could be attacked or threatened with violence" if they refuse to allow their accounts to be used.[1441]

12.20 Cifas reported on money mule recruiters targeting those looking for work, and how they are using the cost-of-living crisis as a tool to enable them to herd victims.[1442] Cifas identified the widely targeted age range for mule activity as 21 to 25,1443 which was the group hardest hit by the economic impact of COVID-19, with thousands facing job losses because of the pandemic, and graduates entering the job market at a time of unprecedented uncertainty.[1444]

[1436] West Yorkshire Police, Money Mules (also known as Squaring). [accessed 5 August 2024].
[1437] FBI Omaha, Money Mule Scheme Targets Teenagers and Young Adults. [accessed 5 August 2024].
[1438] Children's Society response to May 2024 Protection of Children Consultation, p. 11.
[1439] This can include bank account closure, limited access to loans or credit cards, difficulty obtaining a phone contract, and/or a prison sentence of up to 14 years. Source: House of Lords, 2022. Fighting Fraud: Breaking the chain. [accessed 22 September 2023].
[1440] VICE (via YouTube), 2022. The rise of money launderers on Snapchat and Instagram: Crimewave, VICE, 25 October. [accessed 22 September 2023].
[1441] UK Finance, Cifas, n.d. Criminals may ask you to receive money into your bank account and transfer it into another account, keeping some of the cash for yourself. If you let this happen, you're a money mule. You're involved in money laundering, which is a crime. [accessed 22 September 2023].
[1442] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].
[1443] Cifas, 2023. Fraudscape 2023. [accessed 22 September 2023].
[1444] The latest research from Cifas reported 17,157 cases of suspected money muling activity involving 21-30-year-olds in 2020, 5% up on the previous year. This age group accounted for 42% of money mule activity in 2020, up from 38% three years ago. Source: Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].

12.21 Adverse economic conditions may increase the likelihood that people will fall victim, whatever their age or income. Lloyds Bank found that there was a 73% increase in people aged over 40 being used as money mules between August 2023 and July 2024.[1445]

Evidence of risk factors on user-to-user services

12.22 We consider that the risk factors below are likely to increase the risks of harm relating to proceeds of crime. They are also summarised at the start of the chapter.

Risk factors: Service types

12.23 Research indicates that the following types of services can be used to commit or facilitate offences related to the proceeds of crime: social media services, messaging services and gaming services.

Social media services

12.24 Our evidence shows that the risk of proceeds of crime offences taking place on social media services is higher than on many other types of U2U services. Research shows that there has been a rise in money laundering, using popular social media services.[1446] Recruiters of money mules can create user profiles on social media services to join user groups that allow them to find targets,[1447] and use information on users' profiles to befriend potential recruits.[1448] Cifas identifies social media as a "key enabler" in the recruitment of mules.[1449]

12.25 Money mule recruiters have also been known to post fake job opportunities on social media services, which are shown to target young people,[1450] and to use specific terms to attract social media users.[1451]

Messaging services

12.26 Messaging services can also be used in proceeds of crime offences. A money mule recruiter might approach potential money mules through a messaging service or ask people to move onto messaging services after making initial contact on other types,[1452] with research showing that services with robust encryption are increasingly used by money mule recruiters to avoid detection.[1453] [1454]

[1445] Lloyds Bank, 2024. Stubborn as a mule-hunter: Lloyds Bank cracks down on money mules. [accessed 18 October 2024].
[1446] VICE (via YouTube), 2022. The rise of money launderers on Snapchat and Instagram: Crimewave, VICE, 25 October. [accessed 22 September 2023].
[1447] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].
[1448] Cifas, 2021.
[1449] Cifas, 2023. Fraudscape 2023. [accessed 22 September 2023].
[1450] Cifas, 2021.
[1451] Keyworth, M., 2018. I was a teenage 'money mule', *BBC News,* 26 April. [accessed 22 September 2023]. See Risk factors: functionalities and recommender systems section for more information.
[1452] Barclays, n.d. Money Mules: Don't be tricked into committing a crime. [accessed 22 September 2023].
[1453] Cifas, 2021.
[1454] See risks of harm to individuals presented by online proceeds of crime offences for more information.

Gaming services

12.27 Sources also point to gaming services being used by money mule recruiters to first engage with and recruit potential money mules.[1455] [1456] [1457]

Risk factors: User base

User base size

12.28 Social media services with user bases that form large user communities are particularly attractive to bad actors including money mule recruiters[1458] as they make it easy for them to reach large numbers of people[1459] at low cost[1460] and with minimal effort.

12.29 A large user base makes it more likely that posts designed to recruit money mules will reach an even bigger potential audience through user engagement.

12.30 Fraudsters will also make use of or join large, open groups of users, to add authenticity and to look for potential targets. Services with a large user base are more likely to offer such groups.

User base demographics

12.31 The following section outlines significant evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex, and involve multiple factors.

12.32 Data suggests that user base characteristics including age, physical or mental health, media literacy, and socio-economic factors could lead to an increased risk of harm to individuals.

12.33 In relation to money mules, young people may often be targeted by recruiters as their bank accounts can be considered 'clean'. In 2017, Cifas recorded that UK banks had identified 8,500 money mule accounts owned by people under the age of 21 – with some belonging to teenagers as young as 14. Cifas identified 21 to25 years old as the most widely targeted age range for mule activity.1461 This group was the hardest hit by the economic impact of COVID-19, with thousands facing job losses because of the pandemic, and graduates entering the job market at a time of unprecedented uncertainty.[1462] Lloyds Bank reported that students can be particularly vulnerable as they may be seeking extra income, and the Financial Conduct Authority (FCA) found that younger people are the most likely to have

---

[1455] House of Lords, 2022. Fighting Fraud: Breaking the Chain. [accessed 25 September 2023].
[1456] Children's Society Response to Protection of Children Consultation, p. 11.
[1457] FBI Omaha, Money Mule Scheme Targets Teenagers and Young Adults, [accessed 5 August 2024].
[1458] National Crime Agency, 2021. National Strategic Assessment of Serious and Organised Crime. [accessed 22 September 2023].
[1459] Consumers International, 2019. Social Media Scams: Understanding the Consumer Experience to Create a Safer Digital World. [accessed 22 September 2023].
[1460] Federal Trade Commission (Fletcher, E.), 2022. Social Media a gold mine for scammers in 2021. [accessed 22 September 2023].
[1461] Cifas, 2023. Fraudscape 2023. [accessed 22 September 2023].
[1462] The latest research from Cifas reported 17,157 cases of suspected money muling activity involving 21-30-year-olds in 2020, 5% up on the previous year. This age group accounted for 42% of money mule activity in 2020, up from 38% three years ago. Source: Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].

low financial resilience.[1463] [1464] Lloyds Bank also found that "almost one in 10 (9%) of those aged between 18 and 24 years old said they would agree to move money through their bank account in return for a fee or a percentage of the funds"; [1465] this may suggest that young people are less aware of the consequences of engaging in money muling and so are more likely to be herded, or are struggling to find legal alternatives during a difficult economic situation.

12.34 While younger people are the most likely to be involved in money muling[1466] mules are not restricted to this demographic and there have been reported increases in older age groups becoming money mules. This is in part due to the pressures of the cost-of-living crisis and in part, "because larger transactions from their accounts are less likely to arouse suspicions".[1467] [1468]

12.35 Research has shown that in May 2022, 12.9 million UK adults had low financial resilience – nearly one in four (24%) of all UK adults. These people are either already in financial difficulty or could quickly find themselves in difficulty if they suffer a financial shock, as they have little or no savings, or are heavily burdened by their domestic bills or credit commitments. The research also found that people in lower-income households, young adults and those from certain ethnic minorities are more likely to have low resilience or be in financial difficulty.[1469] It may be that they are more likely to be targeted as they may be more likely to be attracted to 'get rich quick' hooks online or are searching content using terms such as 'quick money' which could lead them to recruiters. In adverse economic conditions, the number of people with low financial resilience may grow, increasing the number of people drawn into money muling. There may also be victims who are aware that what they are doing is in some way illegal but do it anyway; the promise of money may outweigh the potential consequences.[1470]

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

12.36 The ability to create a fake user profile provides perpetrators with the opportunity to misrepresent themselves by masking or concealing their official identities. This makes them less traceable and gives them the confidence to operate online and build a criminal network to recruit potential money mules. User profiles operated by perpetrators are often easily

[1463] Lloyds Bank, 2022. Money mules are getting older - with serious penalties for those caught moving scam cash. [accessed 22 September 2023].
[1464] Financial Conduct Authority, 2022. Financial lives 2022 survey: insights on vulnerability and financial resilience relevant to the cost of living. [accessed 22 September 2023].
[1465] Lloyds Bank, 2022.
[1466] Most mules are recruited between the ages of 17 and 24. Source: National Crime Agency, n.d. Young People. [accessed 22 September 2023].
[1467] Hickey, S., 2023. Older people hired as 'money mules' by gangs as cost of living crisis bites, *The Guardian*, 12 June. [accessed 22 September 2023].
[1468] Lloyds Bank, 2022. Money mules are getting older - with serious penalties for those caught moving scam cash. [accessed 22 September 2023].
[1469] Financial Conduct Authority, 2022. Financial lives 2022 survey: insights on vulnerability and financial resilience relevant to the cost of living. [accessed 22 September 2023].
[1470] VICE (via YouTube), 2022. The rise of money launderers on Snapchat and Instagram: Crimewave, VICE, 25 October. [accessed 22 September 2023].

accessible to individuals choosing to make contact and partake in fraudulent activity under the false promise of 'fast money'.[1471]

12.37 Fake user profiles can also serve money mule recruiters, who can create these user profiles to avoid detection. This allows money mules to infiltrate popular groups or special interest pages and find suitable targets on social media services. They will often post images "showing off a luxury lifestyle – for example, expensive cars or large quantities of cash – to entice young people".[1472] These profiles are often designed to attract young people to a luxurious lifestyle.

User profiles

12.38 User profiles, and the information that is often displayed on them, can be used by perpetrators to gather information related to a potential victim. A money mule recruiter will also search a potential victim's user profile, often on a social media service, for information. They can then use this information to befriend a potential money mule or trick them into receiving stolen money in their bank account. This may also happen through a private messaging service.[1473]

User Groups

12.39 UK Finance note that discussion forums and chat rooms are sometimes used to facilitate the offence of arrangements related to criminal property. This includes by "*providing advice or assistance to individuals who are looking to move or conceal assets*."[1474]

User communication

Direct messaging, commenting on content, and re-posting or forwarding

12.40 Direct messaging, re-posting and commenting on posts can allow recruiters to directly contact potential money mules. There is also evidence that recruiters will use specific known terms to attract people, as well as using legitimate job sites on social media to recruit people as money mules.[1475]

Transactions and offers

Post goods or services for sale

12.41 The Crown Prosecution Service notes the existence of 'online marketplaces' used by criminals to sell stolen credit card details, among other items; "These marketplaces are often 'hidden' online, and facilitated by individuals coordinating the trading of these goods".[1476] Online marketplaces typically allow users to post goods or services for sale.

12.42 Money mule recruiters may use job websites and social media services that allow them to post UGC to post fake investment and job opportunities. Research from Cifas and UK

[1471] Bekkers, L.M.J. and Leukfeldt, E.R., 2022. Recruiting money mules on Instagram: a qualitative examination of the online involvement mechanisms of cybercrime, *Deviant Behaviour,* 44(4). [accessed 22 September 2023].
[1472] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].
[1473] Barclays, n.d. Money Mules: Don't be tricked into committing a crime. [accessed 24 August 2023].
[1474] UK Finance Illegal Harms Consultation Response page 4.
[1475] Keyworth, M., 2018. "I was a teenage 'money mule'". *BBC News,* 26 April. [accessed 22 September 2023].
[1476] Crown Prosecution Service, 2019. Cybercrime - prosecution guidance. [accessed 22 September 2023].f

Finance says that this tactic is particularly targeted towards young people whose job prospects have been damaged by the pandemic.[1477]

Content exploring

User-generated content (UGC) searching

12.43 Action Fraud notes that "responding to job adverts, or social media posts that promise large amounts of money for very little work" can put users at risk of becoming a money mule.[1478] Therefore, we consider that being able to find such content proactively through the search function on U2U services would increase the risk that users would encounter and potentially respond to it.

Risk factors: Business models and commercial profile

12.44 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[1477] UK Finance, n.d. Money Mule Recruiters Use Fake Online Job Adverts to Target "Generation Covid?. [accessed 22 September 2023].
[1478] Action Fraud, n.d. Money Muling. [accessed 22 September 2023].

13. Drugs and psychoactive substances

Summary analysis for drugs and psychoactive substances offences: how harm manifests online and risk factors

This chapter summarises the risks of harm from the supply, or offer to supply, drugs and psychoactive substances.

The harm to individuals resulting from the supply and use of drugs and psychoactive substances is substantial, and online services play an important role in facilitating the supply and distribution of illegal drugs in the UK. In 2022 there were 4,907 deaths related to drug poisoning registered in England and Wales, the highest number since records began in 1993, and 1.0% higher than in 2021. Under-18-year-olds may be particularly at risk of the harms brought about by illicit substances, and young people are most likely to be exposed to the related risks online. Surveys show that a wide range of services are being used to access drugs, with social media services and some video-sharing services being the most used. Law enforcement reported in 2024 that 20 to 24% of teenagers report having seen drugs for sale on social media. The impact of the trade and use of illicit substances is felt keenly on society, not just through associated criminality but also the burden it places on health and other public services.

Service type risk factors:

From our research, there is significant evidence supporting the role of **social media and video sharing services** in facilitating or committing drugs and other psychoactive substance offences. Suppliers use these services to advertise the sale of these illicit substances and then use **messaging services** to negotiate transactions. Our research identifies three additional service types that can also be linked to the supply of drugs and other psychoactive substances; however, the known use of these services is much more limited: **discussion forums and chat room services, dating services, marketplace and listing services**.

User base risk factors:

Under-18s tend to receive more content promoting drugs on certain social media services than over-18s and therefore our research has identified **age** and potentially **gender** as a risk factor. Gender is not necessarily a defining factor for users who sell and purchase drugs and psychoactive substances online. However, men tended to see more Class A drugs promoted online than women.

Functionalities and recommender systems risk factors:

The ability to **post content,** in particular, images and emojis, as well as **tag content** have been identified as important functionalities in the supply of drugs and

psychoactive substances online, as these can be used to promote drugs and signpost potential buyers. This can all be supported by the ability to create **anonymous user profiles** which provides privacy to users conducting illegal activity. Similarly, **direct messaging** allows suppliers to talk with their customers away from larger user groups. Although it is often **encrypted messaging** that is favoured over messaging without automatically enabled encryption, due to its added security. **Ephemeral messaging** can also encourage perpetration by limiting digital traces of purchases.

Network recommender systems can recommend other dealers to users and allow them to increase their exposure. This risk can be amplified depending on the design of the recommender system used by the service. Likewise, **hyperlinking and user-generated content (UGC) searching** can be popular methodologies for linking users to additional illicit content. Users often have to connect with potential dealers through **user connections** before viewing their **user profiles** and associated content. Menus or images depicting the products for sale can be posted on services, or in closed **user groups**.

It is also possible to link **posting goods or services for sale, reacting to content and commenting on content** to the buying and selling offenses connected to drugs and psychoactive substances. However, there is limited research in this chapter to identify their role as a key facilitator of this harm.

Introduction

13.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the drugs and psychoactive substances offences listed under 'Relevant offences'; and
- the use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

13.2 We set out the characteristics of U2U services that we consider are likely liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

13.3 Although we have not considered specific evidence about the dark web for the purpose of this chapter, we acknowledge it continues to play a role in the online drug supply market.[1479] [1480]

[1479] Commission On Combating Synthetic Opioid Trafficking, 2022. Commission on Combating Synthetic Opioid Trafficking: Technical Annexes [Accessed 17 October 2022].
[1480] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA (Demant, J. and Bakken, S.A.), 2019. Technology-facilitated drug dealing via social media in the Nordic countries. [accessed 17 October 2022], p. 15. [Accessed 17 October 2022].

Relevant offences

13.4 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding the supply of drugs and psychoactive substances, Ofcom is required to consider the risks of harm connected with the following priority offences under Schedule 7 of the Act:[1481]

- Unlawful supply, or offer to supply, of controlled drugs[1482]
- Prohibition of supply of articles for administering or preparing controlled drugs[1483]
- Inciting any offence under the Misuse of Drugs Act[1484]
- Supplying, or offering to supply, a psychoactive substance.[1485]

13.5 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences (and in relation to offences in Scotland, being involved art and part in the commission of these offences).

13.6 For more details on the offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance or ICJG.

How drugs and psychoactive substances offences manifest online

13.7 This section is an overview that looks at how drugs and psychoactive substances offences manifest online, and how individuals may be at risk of harm.

13.8 The supply of drugs and psychoactive substances is facilitated online by enabling suppliers to market their products and connect with potential buyers. Many of the functionalities of online services that enable suppliers of legal products to reach, engage and deal directly or indirectly with customers can also be used by those selling illicit substances. Moreover, online modes of drug provision are perceived to offer benefits to both sellers and buyers compared to traditional face-to-face drug dealing, such as a greater variety of substances, greater accessibility of illegal substances, and less exposure to police detection.[1486]

13.9 To put the risks of harm into context, law enforcement reported in 2024 that 20 to 24% of teenagers have seen drugs for sale on social media.[1487] In 2021, 35% of 13–18-year-olds surveyed by the Daniel Spargo-Mabbs Foundation reported having seen illegal drugs for

[1481] As per the Misuse of Drugs Act 1971, 'Supplying' includes distributing, and 'Controlled drugs' refers to the definition listed in Schedule 2 and categorised as Class A (e.g., cocaine, ecstasy), Class B (e.g. cannabis, codeine) and Class C Drugs (e.g. benzodiazepines, diazepam). A 'psychoactive substance' is defined as a substance which is capable of producing a psychoactive effect on a person who consumes it.
[1482] Section 4(3) of the Misuse of Drugs Act 1971.
[1483] Section 9A of the Misuse of Drugs Act 1971. There is very limited evidence linked to this particular offence; for further information on this offence, please refer to the Illegal Content Judgements Guidance. Throughout this chapter, we expect the risk factors associated with this offence to be largely similar to the offence of unlawful supply, or offer to supply, of controlled drugs.
[1484] Section 19 of the Misuse of Drugs Act 1971.
[1485] Section 5 of the Psychoactive Substances Act 2016.
[1486] Dewey, M. & Buzzetti, A. 2024. Easier, faster and safer: The social organization of drug dealing through encrypted messaging apps. *Sociology Compass, 18(2)*. [accessed 22 October 2024].
[1487] National Crime Agency (NCA), 2023.

sale on social media sites or apps.[1488] A 2019 Volteface poll of 16–24-year-olds found that one in four young people had seen illicit drugs advertised for sale on social media. Out of those who reported seeing illicit drugs for sale on social media, Cannabis was identified as the most seen drug, with nearly two-thirds (63%) of respondents saying they had seen it offered for sale on social media services, followed by cocaine (26%) and MDMA (24%).[1489] Among those aged under 18 the figures cannabis was still the most common drug seen (70%), followed by cocaine (28%) and MDMA (35%).[1490]

Risks of harm presented by supply or offer to supply controlled drugs and psychoactive substances

13.10 In 2022, it was estimated that one in 11 adults (9.2%) aged 16 to 59 in England and Wales had taken a drug in the past year, with use higher among 16 to 24-year-olds (18.6%).[1491] Cannabis, cocaine and ecstasy, were among the main drugs misused by young people who required substance misuse treatment from 2020 to 2021.[1492] In 2022 there were 4,907 deaths related to drug poisoning registered in England and Wales, which according to the ONS is the highest number since records began in 1993 and is 1.0% higher than in 2021.[1493] The financial impact of drug use and drug misuse costs over the UK over £20 billion per year.[1494] While it cannot be assumed that these drugs were purchased online, research indicates that online mediums will have facilitated in the procession and sale of the illicit substances.

13.11 Evidence of the risks of harm arising specifically from the supply of drugs and psychoactive substances online is relatively limited. However, it is likely that this offence leads to a similar experience as for those who use drugs obtained by more traditional means.

13.12 This could include significant health risks, such as the development of drug use disorders. Drug use disorders increase morbidity and mortality risks, and can lead individuals to suffer from issues in their personal, family, social, educational, and occupational relationships.[1495] The effect of illicit drug supply is far wider than for those that consume them and include societal concerns like the impact on mental health and the burden on the National Health

[1488] Please note the survey sample of 1,919 young people was self-selecting so it is not possible to determine whether it was representative of the whole population of 13–18-year-olds in the UK. Source: Daniel Spargo-Mabbs Foundation, 2021. Young people, drugs and social media – a survey for 13-18 year olds. [accessed 31 October 2024].
[1489] Volteface, 2019. DM for details: Selling drugs in the age of social media [Accessed 17 October 2022].
[1490] Volteface., 2019.
[1491] Office for National Statistics, 2022. Drug misuse in England and Wales: year ending June 2022. [accessed 7 August 2023].
[1492] In addition, a significant proportion of young people reported a problem with alcohol (41%) and nicotine (12%): Office for Health Improvement and Disparities, 2022. Young people's substance misuse treatment statistics 2020 to 2021: report. [accessed 8 May 2023].
[1493] Office for National Statistics, 2022. Deaths related to drug poisoning in England and Wales: 2021 registrations. [accessed 8 May 2023].
[1494] UK Parliament, 2024. Reducing the harm from illegal drugs [accessed 30 October 2024].
[1495] World Health Organisation, 2022. Drugs (psychoactive). [accessed 24 October 2022].

Service.[1496] [1497] The risks associated with traditional street dealing also continue to be prevalent as a result of the options for in-person collections.[1498]

13.13 Under-18-year-olds may be particularly at risk of the harms brought about by illicit substances. Evidence indicates that the use of drugs when an individual's brain is not yet fully developed, can be linked to an increased risk of the onset of depression and suicidal tendencies.[1499] [1500]

13.14 Young people can often be susceptible to exploitation in the supply of drugs. Barnardo's, a UK charity focused on vulnerable children, highlights that among other forms of exploitation, children and young people "*are coerced to carry drugs and weapons from one area to another to service complex drug supply chains.*"[1501]

13.15 Another concern presented by the online illegal drug market is the risk associated with counterfeit drugs. Substandard, spurious, falsely labelled, falsified, and counterfeit (SSFFC) medical products are an increasing threat to consumer health and are often associated with nitazenes.[1502] [1503] Since 2009, a total of 81 new opioids have been identified on the European drug market with seven new substances notified in 2023, six of which were nitazene opioids.[1504] It is highly likely that, when users are purchasing drugs that are mislabelled, such as counterfeit oxycodone and benzodiazepine tablets the user is unaware that they are consuming a nitazene tablet. This heightens the risk to the user due to the potency and unwanted effects of nitazenes.[1505] In the UK, between 01 June 2023 and 15 August 2024, there were 284 confirmed fatalities alongside a number of confirmed or suspected near fatal overdoses involving nitazenes.[1506] The World Health Organisation

---

[1496] Between 2019/2020 there were "7,027 hospital admissions for drug-related mental and behavioural disorders", "16,994 hospital admissions for poisoning by drug misuse" and "99,782 admission with a primary or secondary diagnosis of drug-related mental and behavioural disorders": NHS England, 2021. Statistics on Drug Misuse, England 2020. [accessed 18 November 2024].

[1497] In 2022 the Government published statistics relating to the Children in Need Census; in the year up to the 31 March 2022, Drug misuse concerns relating to the child and a parent were a factor in 92,250 cases. 'Children in Need' are a legally defined group of children (under the Children Act 1989), assessed as needing help and protection as a result of risks to their development or health. The latest data for 'Children in Need' is available: UK Department of Education, 2024. Reporting Year 2024. [accessed 18 November 2024].

[1498] Debt bondage, a real or perceived debt used as a method to exert control over individuals to carry out tasks including drug dealing, is a common risk associated with street-level drug dealing. Such debt can be incurred through accepting drugs as a 'gift', which recipients are then expected to repay: Crown Prosecution Service, 2022. County Lines Offending [accessed 18 November 2024]; Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, *International Journal of Drug Policy*, 63, 101-110. [accessed 3 June 2023]; Volteface, 2019.

[1499] Gobbi, G., Atkin, T. and Zytynski, T. 2019. Association of Cannabis Use in Adolescence and Risk of Depression, Anxiety, and Suicidality in Young Adulthood AMA Psychiatry: A Systematic Review and Meta-analysis. JAMA Psychiatry, 76(4):426-43. [accessed 24 October 2024].

[1500] Kiburi, S. K., Molebatsi, K., Ntlantsana, V. and Lynskey, M. T. 2021. Cannabis use in adolescence and risk of psychosis: Are there factors that moderate this relationship? A systematic review and meta-analysis. *Substance Abuse*, 2021;42(4):527-542. [accessed 24 October 2024].

[1501] Barnardos, 2023. Child exploitation: a hidden crisis. [accessed 28 October 2024].

[1502] A Nitazene is a Class A synthetic opioid associated with mimicking the effects of natural opioids. These drugs have a high potency and are often cut with additional drugs increasing the risk of overdose.

[1503] An example is Alprazolam, a medicine in the benzodiazepine family of drugs. This is ten times stronger than diazepam and is not prescribed by the National Health Service. The UK Government has recognised that Alprazolam, in the form of counterfeit Xanax tablets, is a commodity available in street-level drug dealing markets and on illegal website and social media service: UK Health Security Agency, 2024. Alprazolam (Xanax): What are the facts? [accessed 24 October 2024].

[1504] One in 2022 and six in 2021. Source: NCA, 2024.

[1505] NCA, 2024.

[1506] NCA, 2024.

(WHO) recognises counterfeit drugs as one of the urgent health challenges for the next generation.[1507]

13.16 The online sales of drugs can also be linked to other criminality facilitated by the use of online services. Law enforcement identify that Chemsex-context illicit substances,[1508] including methamphetamines and GHB/GBL, are often associated with sexual crime, such as the illicit recording and online sharing of sexual activity.[1509]

Evidence of risk factors on user-to-user services

13.17 We consider that the risk factors we've listed are likely to increase the risks of harm relating to the sale or supply of drugs and psychoactive substances. These are also summarised at the start of this chapter.

Risk factors: Service types

13.18 Research indicates that the following types of services can be used to facilitate or commit offences related to the sale of drugs and psychoactive substances: social media services, video-sharing services, and private messaging services.

Social media services and video-sharing services

13.19 Researchers highlight the importance of specific functionalities (discussed further in this chapter), and the apparent effectiveness of service provider's moderation processes, on whether content that is likely to amount to drugs and psychoactive substances offences is directly accessible to users and researchers conducting online investigations.[1510] However, a wide variety of sources indicate that the offer to supply drugs and psychoactive substances manifests to a significant extent on social media services. Fundamentally, this is because these services can allow buyers and sellers to connect – posting information about where to obtain drugs and psychoactive substances as well as posting items for supply, and then facilitating further engagement to conclude purchases.

13.20 A study conducted in the UK by the Daniel Spargo-Mabbs Foundation in 2021 found that 25% of young people aged 13-15 reported having seen illicit drugs advertised on social media services.[1511] Similarly, a survey conducted by Moyle *et al*. showed that a wide range of services were reported as being used to access drugs, with social media services and some video-sharing services being the most used.[1512]

[1507] World Health Organization, n.d. Substandard and falsified medical products. [accessed 15 October 2024].
[1508] Chemsex-context illicit substances are drugs that are associated with sexual activity. The relationship between these drugs and sexual behaviour, creates a link with harms of a sexual nature.
[1509] NCA response to the November 2023 Illegal Harms Consultation.
[1510] For instance, researchers in Europe highlighted the relative ease with which they identified suspected drug dealers on Snapchat compared to Facebook and Instagram using the same search queries and slang terminology European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (Demant, J. and Aagesen, K.M.B.), 2022. An analysis of drug dealing via social media. [accessed 22 October 2024].
[1511] The Independent. 2021. One in five 13-14-year-olds have seen drugs being sold on social media [accessed 25 October 2024].
[1512] Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, *International Journal of Drug Policy*, 63, 101-110. [accessed 3 June 2023] [accessed 17 October 2022].

13.21 Studies looking at the content of drug posts on social media services have found significant numbers of posts, but there are conflicting conclusions on how much of this content relates to illegal drug supply. For example, some evidence shows that there has been a large number of posts on Instagram that advertise the sale of drugs.[1513] Another study found a large number of Fentanyl-related posts identified on Twitter of which a very small sample were determined to be promoting the marketing and sale of Fentanyl.[1514] A 2023 multidisciplinary scoping review found that, on average, across 56 peer-reviewed studies on the sale of drugs online, for every 100 social media posts related to drugs or drug-selling behaviour researchers assessed, approximately 13 appeared to offer illicit drugs for sale.[1515]

13.22 Similarly, a study analysing posts on Instagram related to various controlled substances and illicit drugs indicated that of the many posts related to them, there were far fewer posts that explicitly included an offer for supply or an offer to purchase the substances.[1516] Still, the authors concluded that "*users have active conversations about selling and buying drugs, meaning that these social media posts act as digital marketplaces for drug dealing*."

13.23 A recent study on image search for Fentanyl-related precursors[1517] found that a large number of image search results were sourced from Pinterest (600+ URLs), followed by Facebook (200 to 300 URLs), LinkedIn (200+ URLs), Twitter (100+ URLs) and Tumblr (0 to 100 URLs).[1518] From this study we can assume that pictures of drugs are being posted on social media services, and it is possible some of this content might be related to the supply of these drugs.

13.24 Functionalities typically present on social media services and video-sharing services can be used in the supply of drugs and psychoactive substances. These include user groups and user connections[1519] which can be used to connect, network, and establish trust between dealers and potential buyers. Posting content on social media services and video-sharing services can also help dealers advertise and sell drugs.[1520]

[1513] Petersen et al. found that of the 152,308 pieces of online content identified specifically relating to study drugs on Instagram, 27.3% related to illicit drugs supply. The majority of this content also emphasised the benefits of using these drugs. Petersen et al. 2021. #studydrugs-Persuasive posting on Instagram [accessed 17 October 2022].

[1514] Mackey and Kalyanam found that of the 28,711 Fentanyl-related posts identified on Twitter during 2015, a period when the Fentanyl crisis was escalating, only 771 (<1% of total) were determined to be promoting the marketing and sale of Fentanyl and other controlled substances online after isolating posts relating to news reports. Mackey and Kalyanam. 2017. Detection of illicit online sales of fentanyls via Twitter [accessed 17 October 2022].

[1515] Fuller, A., Vasek, M., Mariconti, E., and Johnson, S.D., 2023. Understanding and preventing the advertisement and sale of illicit drugs to young people through social media: A multidisciplinary scoping review. [accessed 21 June 2024].

[1516] A 2018 study collected a total of 12,857 posts from Instagram relating to Xanax, OxyContin, LSD and MDMA. They found a total of 1,228 posts by those likely to be drug dealers, comprising 267 unique users. Of the 1,228 detected posts analysed, 232 explicitly included an offer for supply or an offer to purchase. These included posts, or comments within posts, from users offering to supply drug(s) (with contact information), and comments from other users asking for more information or requesting to buy the drug. Li, J., Zu, Q., Shah, N. and Mackey, T. 2018. A Machine Learning Approach for the Detection and Characterization of Illicit Drug Dealers on Instagram: Model Evaluation Study [accessed 17 October 2022]

[1517] A precursor is a chemical needed to synthesise a drug.

[1518] Commission On Combating Synthetic Opioid Trafficking, 2022. Commission on Combating Synthetic Opioid Trafficking: Technical Annexes E-14. [accessed 17 October 2022].

[1519] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022]. See functionalities section for more information.

[1520] Volteface, 2019.

Messaging services

13.25 Messaging services are commonly used to offer and facilitate the supply of drugs, as they offer a closed channel of communication that reduces the risk of detection for buyers and suppliers. This is supported by various studies which show that suppliers often redirect prospective buyers to private messaging services, particularly those with encryption, to carry out negotiations and transactions over closed channels of communication.[1521]

13.26 Direct messaging, a functionality that is central to private messaging services, as well as encrypted messaging were found to be particularly important risk factors, providing a level of privacy and security to both sellers and buyers.[1522] Some messaging services also have wider functionality, enabling the building of communities and even e-commerce functionality - this makes such encrypted messaging services effective places to facilitate the sale of drugs.[1523] The latter supports the conclusion that private messaging services with encryption are an important service type used in the supply of drugs and psychoactive substances.[1524]

Discussion forums and chat room services

13.27 Large discussion forums have been identified in research exploring online drug selling networks. Research in Denmark and Sweden found "*vast amounts*" of drug dealing-related content on online discussion forums, including fora that appeared to be created and moderated by drug dealers.[1525] These findings are not limited to the Nordic countries the research focused on, and there are also examples from the UK.

13.28 Law enforcement investigations have identified instances of the advertised sale of drugs in message forums and interest sites where the main topic is unrelated to drugs or illicit substances or activities.[1526] Sellers often offered multiple illicit substances for sale in the same advert.[1527]

Dating services

13.29 Dating sites have also been linked by law enforcement to the sale of drugs online. Drugs adverts on some dating and hookup apps, particularly those apps which cater to narrow audiences, are often more subtle than those on open social media, using more slang terminology, with the recipient audience assumed to be more likely to have a greater level

---

[1521] Volteface, 2019; Moyle, L., Childs, A., Coomer, R., and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, International Journal of Drug Policy, 63 101-110 [accessed 3rd June 2023].

[1521] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (Demant, J. and Bakken, S.A.), 2019. Technology-facilitated drug dealing via social media in the Nordic countries. [accessed 17 October 2022].

[1522] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (Demant, J. and Aagesen, K.M.B.), 2022. An analysis of drug dealing via social media. [accessed 22 October 2024].; EMCDDA (Demant, J. and Bakken, S.A.), 2019: Volteface, 2019.

[1523] Dewey, M. & Buzzetti, A. 2024. Easier, faster and safer: The social organization of drug dealing through encrypted messaging apps. *Sociology Compass, 18(2)*. [accessed 22 October 2024].

[1524] Volteface, 2019. DM for details: Selling drugs in the age of social media [Accessed 17 October 2022]; Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, *International Journal of Drug Policy*, 63, 101-110. [accessed 3 June 2023]; C4ADS. 2020. Lethal Exchange: Synthetic Drug Networks in the Digital Era [accessed 17 October 2022].

[1525] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (Demant, J. and Aagesen, K.M.B.), 2022. An analysis of drug dealing via social media. [accessed 22 October 2024].

[1526] Interest sites refers to websites dedicated to a particular subject such as gaming, fashion and motoring.

[1527] NCA response to the November 2023 Illegal Harms Consultation.

of knowledge of drugs purchase options.[1528] High-risk high-harm substances, such as methamphetamine, GHB/GBL, and synthetics, are often advertised for sale as 'bundles' of multiple substances on dating apps. This is of particular concern to law enforcement as the increase in poly-drug use is believed to be a factor linked to the increase in drug-related deaths across the UK.[1529]

Marketplaces and listing services

13.30 Law enforcement have identified that there are an estimated 35,000 to 40,000 e-commerce websites selling pharmaceutical drugs globally, of which 95% operate illegally by selling prescription, controlled, counterfeit, substandard or out-of-date drugs.[1530] In the UK, over 3 million medicines and medical devices valued at over GBP 9 million were seized in 2021 and 113,000 illegally operating websites removed. With this knowledge we can suspect that there is risk of harm occurring on Marketplace and listing services under the false pretence of legal operations. This has also been supported by the International Narcotics Control Board Annual reports which identified how e-commerce platforms were used in the emerging illicit substances.[1531]

Risk factors: User base

User base demographics

13.31 The following section outlines key evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

13.32 Data suggests that user-base characteristics including **age** and potentially **gender** can lead to increased risks of harm.

13.33 Our analysis suggests that children are particularly vulnerable to harm from the supply and offer of illegal drugs online. Evidence indicates that young people can be groomed on social media services for *"roles within the drug supply chain"*.[1532] Other evidence found under-18-year-olds were as likely or more likely to see drugs promoted on specific U2U services than over-18s.[1533]

13.34 Law enforcement investigations have identified that drugs are often packaged in similar ways to confectionery or sweets popular with children, raising concerns that this reduces some barriers to purchase and increases the risk to children. Examples of this can include Tetrahydrocannabinol (THC) edibles in the form of gummy sweets and brightly coloured THC vapes.[1534] By packaging these drugs in misconceiving child friendly packaging, children are likely to be mislead into purchasing these illegal products.

[1528] NCA, 2023.
[1529] NCA response to the November 2023 Illegal Harms Consultation.
[1530] NCA, 2023.
[1531] International Narcotics Control Board, 2023. Annual Report 2023. [accessed 15 October 2024].
[1532] Crest Advisory (Caluori, J., Mooney, B. and Kirk, E.), 2023. Running out of credit: Mobile phone tech and the birth of county lines. [accessed 12 May 2023].
[1533] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].
[1534] NCA response to the November 2023 Illegal Harms Consultation.

13.35 Gender cannot be established as a risk factor; however, there is some research suggesting that more men than women reported seeing crack cocaine and heroin advertised for sale online,[1535] while more women than men had seen Xanax advertised for sale (23% versus 18%).[1536]

Risk factors: functionalities and recommender systems

User identification

User profiles

13.36 Volteface research found that user profile features were used to promote drugs for sale.[1537] According to this report, it was common for suspected drug dealer accounts to have photos and/or videos of their drugs on their user profiles, particularly those that sold cannabis.[1538]

13.37 Moyle *et al.* found that the biography feature on a user profile was used by all the suspected dealers the researchers found in their research, often providing a useful indication of whether that account was involved in illicit drug supply or not.[1539]

13.38 The Volteface study also found that the biography feature on a user profile was used by all the suspected dealers that the researchers found, often providing a useful indication of whether or not that account was involved in illicit drug supply. The study additionally identified that the biography or 'intro' section of user profiles on social media services can be used by suspected dealers to direct users to some private messaging services.[1540]

Anonymous User profiles

13.39 Research has found that anonymous user profiles are instrumental for users suspected of illegal drug dealing. The profile images are often unidentifiable or anonymous, usually showing a drug-related image and multiple accounts, usually with similar names on their profile.[1541] Evidence shows that the trade of illicit substances can be supported by anonymity [1542] and that large networks of anonymous illicit activity can even work to create a safer environment for illegal activity by reducing the exposure to additional harms (for example, violence) that can often be associated with offline markets.[1543]

[1535] Crack cocaine seen advertised online for sale on social media by 16 to 24 year olds: 16.3% male vs 7.7% female. Heroin seen advertised online for sale on social media by 16 to 24 year olds: 10.3% male vs 4.7% female. Source: Volteface, 2019.
[1536] Volteface, 2019.
[1537] Volteface, 2019.
[1538] Volteface, 2019.
[1539] Examples include users including emojis in their user profiles that other users could look for to identify a potential dealer. Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, *International Journal of Drug Policy*, 63, 101-110. [accessed 3 June 2023].
[1540] Volteface, 2019.
[1541] "*For example, one account had two profiles where the only difference was the surname: 'Green' and 'Greenn'".* Source: Volteface, 2019.
[1542] Hammond, A.S., Paul, M.J., Hobelmann, J., Koratana, A.R., Dredze, M. and Chisolm, M. 2018. Perceived Attitudes About Substance Use in Anonymous Social Media Posts Near College Campuses: Observational Study. *JMIR Mental Health* 5(3). [accessed 17 October 2022].
[1543] Sanden, R.v.d., Wilkins, C. Rychert, M. and Barratt, M.J., 2022. The Use of Discord Servers to Buy and Sell Drugs, *Contemporary Drug Problems*, 49(4), 453-477. [accessed 18 November 2024].

User networking

User connections

13.40 User connections are a risk factor in the context of the offences described in this chapter. Many drug dealers' user profiles are closed to the public, so that prospective customers must request to connect before they can see the user profile, user connections and content. It is common for suspected dealers to connect with each other, which may provide another way of finding users offering drugs for sale.[1544]

13.41 The Volteface report found that dealers often posted heavily with the aim of getting potential customers to notice them and potentially 'follow' them back.[1545]

User groups

13.42 Setting up user groups has been shown to facilitate relevant offences. Services enable customers to join any public groups in relation to their topic of interest. There are a large range of groups relating to online drugs markets. These vary in terms of the type of drug, geographical location, or whether the transaction is face-to-face or postal. Customers then progress from public groups to private groups via invite links.[1546] Likewise, closed user groups on some social media services, where an invitation is required, allow suppliers to promote their services, providing a contact for potential buyers.

13.43 Analysis by the Centre for Advanced Defence Studies found that users who are seeking to sell and purchase illicit substances come together in private user groups on a social media service. Users discuss relevant drug laws, products are advertised or reviewed, and buyers alert other users to potential seller scams. Such scams include 'sellers' who fail to supply the product after receiving payment.[1547]

13.44 A 2019 study by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), which focused on Iceland, Sweden, Denmark, Norway and Finland, found that Denmark, Iceland and Sweden have an active open social media drug market, noting the role of user groups on one social media service in particular. The use of closed user groups in Norway was also highlighted.[1548]

User communication

Direct messaging

13.45 Direct messaging is a functionality used to offer to supply drugs. For example, the same EMCDDA study found that suppliers and buyers used direct messaging on services. Access therefore requires a high-level of previous knowledge to be able to contact a seller, such as knowing who to contact and how.[1549]

13.46 Volfteface found that drug dealers promote drugs on their user profiles or other more open forms of communication like captions to photos or videos, which would often instruct individuals that they should "*direct message or dm them".* Researchers concluded that

---

[1544] Volteface, 2019.
[1545] Volteface, 2019.
[1546] NCA, 2024.
[1547] C4ADS, 2020. LETHAL EXCHANGE: SYNTHETIC DRUG NETWORKS IN THE DIGITAL ERA. [accessed 17 October 2022].
[1548] European Monitoring Centre for Drugs and Drug Addiction (Demant, J. and Bakken, S.A.), 2019. Technology-facilitated drug dealing via social media in the Nordic countries. [accessed 17 October 2022].
[1549] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (Demant, J. and Bakken, S.A.), 2019.

"*communication between potential customers and drug dealers would happen privately via the direct messaging function*".[1550]

13.47 The Moyle *et al*. 2019 report found a specific social media service to be one of the most popular for purchasing and/or selling drugs and other psychoactive substances. Messages were sent from suppliers to their connections to offer illicit drugs for sale. Text within ephemeral messages would describe the drugs and how to obtain them, with emojis occasionally used in place of text. Users were then able to message through the service or continue the conversation elsewhere, as indicated by the supplier.[1551]

Encrypted messaging

13.48 Encrypted messaging is an essential component in the supply of drugs and psychoactive substances. Open adverts commonly redirect the customer to encrypted messaging apps or closed groups for more details, sales or enquiries, with new subscribers manually approved and moderated by the group owner.[1552] While potential perpetrators have been shown to use direct messaging for the sale of drugs and psychoactive substances, direct messaging that offers end-to-end encryption is particularly risky due to the added security it offers. Perhaps for this reason, dealers appear to favour the use of private messaging services that have automatically integrated encrypted messaging.

13.49 For example, Facebook Messenger provides a secure messaging service which dealers and buyers use to arrange deals with known suppliers.[1553] But the evidence also showed that it was unusual to see dealers signposting users to contact them on Facebook Messenger. Volteface hypothesises that this may be because at the time, this service only used end-to-end encryption if the user turned it on manually. It was more usual for dealers to direct potential buyers to use private messaging services which automatically provided end-to-end encryption, or to phone or text them privately.[1554]

13.50 With regards to direct messaging, Volteface concluded that "*it was common for dealers to navigate potential customers to alternative encrypted methods of communication*".[1555]Analysis from the Centre for Advanced Defence Studies also demonstrated that suppliers of synthetic drugs use private Facebook groups to establish buyers' trust and often suggest continuing purchase conversations on private messaging services which provide end-to-end encryption.[1556]

Ephemeral messaging

13.51 There is less evidence that drug dealers can reach customers through ephemeral messaging, which may be due to its nature. However, Volteface research highlights the use of 'stories', which are ephemeral and often visible for 24 hours on several social media and video sharing services. They are used to communicate details of the sale relating to their

---

[1550] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].
[1551] Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, *International Journal of Drug Policy*, 63, 101-110. [accessed 3 June 2023].
[1552] NCA, 2024.
[1553] Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs. [accessed 17 October 2022].
[1554] Volteface, 2019. DM for details: Selling drugs in the age of social media [Accessed 17 October 2022].
[1555] Volteface, 2019.
[1556] C4ADS, 2020. Lethal Exchange: Synthetic Drug Networks in the Digital Era [Accessed 17 October 2022].

offer to supply illicit drugs.[1557] While stories are not messages and are more closely related to the posting of content, they appear to be used in part because they are time limited.

13.52 Services with ephemeral messaging have auto-destruction or 'burn on read' settings users can apply to their messages. This may reassure some users that their digital trace is concealed, which appears to have some traction in online communities. Here, users contrast the insecurity of text messages and phone calls with the 'safety' of the service, which they assume does not store a data base of users' photos, videos and text. Survey and interview data suggest that suppliers often advertise their products on these services and then require buyers to close the deal on encrypted services.[1558] Law enforcement have suggested that a video-sharing service's use of 'ephemeral messaging' is a common external method of contact linked by dealers in their adverts on services.[1559]

Reacting and commenting on content

13.53 Content reactions such as 'likes' and comments on user profiles act as reviews on online marketplaces and can also provide a sense of security regarding the reliability of a particular seller.[1560]

Posting content (images, videos, text, emojis)

13.54 The ability to post content, in particular images and emojis, is an important functionality in the supply of drugs and psychoactive substances online, as it can be used to promote drugs and signpost potential buyers. Studies have found that dealers can be open about the supply of drugs in the content they post; for example, with delivery status updates or pictures clearly showing the drugs. In this content, dealers can tell potential buyers to contact them via a direct message, where they can provide a price list or menu of the drugs they sell.[1561] Dealers also caption pictures of drugs with the product names.[1562] Additionally, Dealers were found to post images of block text rather than pictures of the product to promote their produce and prices.[1563]

13.55 An EMCDDA study that analysed individual posts within social media groups involved in drug-dealing activity found that posts offering drugs dominated the groups' activities, with 50% of the collected posts offering drugs for sale.[1564]

13.56 Volteface found that it was *"common for dealers to post their drug 'menus' and price lists in their stories"*. This would include what drugs were available that day, with quantities and prices, and sometimes phone numbers so that customers could get in direct contact with the dealer.[1565]

[1557] Volteface, 2019..
[1558] Moyle et al. 2019.
[1559] NCA response to the November 2023 Illegal Harms Consultation.
[1560] Moyle et al. 2019.
[1561] Volteface, 2019. DM for details: Selling drugs in the age of social media [Accessed 17 October 2022]
[1562] "For example, to indicate the strain of cannabis, the account would describe it as 'Lemon Haze'". Source: Volteface, 2019.
[1563] For example: "200 vals £90 200 lorazepam £120". Source: Volteface, 2019.
[1564] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (Demant, J. and Bakken, S.A.), 2019. Technology-facilitated drug dealing via social media in the Nordic countries [Accessed 17 October 2022].
[1565] Volteface, 2019.

13.57 The images and videos posted by suspected dealers are often captioned and make use of emojis instead of words in their posts offering drugs, to help avoid content moderation.[1566] Although images of potentially harmful content can be recognised by content moderation tools, the multiple formats these drugs can take make it more difficult to identify illegal substances.[1567]

13.58 Dealers usually post frequently about their activity (known colloquially as 'dealers spam'), posting multiple videos and a range of images of advertised products to followers on social media services. In some circumstances, dealers would, for example, send out several messages a day to say what products they had, and to notify of any special offers. Dealers would also 'prove' the quality and legitimacy of their product by posting videos of themselves using the products.[1568]

Transactions and offers

Posting goods or services for sale

13.59 Evidence indicates that the ability to post goods or services for sale can be used by suppliers to offer and sell drugs online. The Volteface study found that 72% of young people see illegal drugs offered or 'advertised' "*on social media services or apps at least once a month or more*". The study also found evidence of suspected drug dealers using functions whereby users can post an item for sale. For instance, the study noted that "*Facebook also offers a function whereby users can post an item for sale and the researchers saw evidence of this being used by suspected drug dealers*."[1569]

Content exploring

User-generated content searching

13.60 Searching for drug supply on social media services may be an effective way for users to find illicit drugs and psychoactive substances. Volteface found that *"searching on social media sites or apps"* ranked second of the four options[1570] when survey respondents were asked which method was easiest for obtaining contact details for a drug dealer.[1571]

Content tagging

13.61 Tagging and labelling content has been found to be an effective way for dealers to consolidate and drive potential buyers to their supply. Hashtags can be used to broaden reach to potential buyers.[1572]

13.62 The Commission on Combating Synthetic Opioid Trafficking found that most content promoting Fentanyl on Pinterest was given misleading labels by the author. This was

---

[1566] Volteface, 2019.
[1567] National Crime Agency response to the November 2023 Illegal Harms Consultation
[1568] Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, *International Journal of Drug Policy*, 63, 101-110. [accessed 3 June 2023].
[1569] Volteface, 2019.
[1570] The other options were asking a friend, asking a family member, and asking a stranger. Source: Volteface, 2019.
[1571] Volteface, 2019.
[1572] For example, #buy, #sell, #buypainmeds, #drugsforsale, #opioids, #painmeds, and #controlled. Source: Mackey, T., Kalyanam, J., Klugman, J., Kuzmenko, E and Gupta, R. 2018. Solution to Detect, Classify, and Report Illicit Online Marketing and Sales of Controlled Substances via Twitter: Using Machine Learning and Web Forensics to Combat Digital Opioid Access. *Journal of Medical Internet Research*, 20(4). [accessed 18 November 2024]

understood to be a method for the authors to circumvent automated content moderation.[1573]

13.63 Hashtags are also used to search for potential sellers.[1574]

Hyperlinking

13.64 It is widely documented that suspected dealers redirect prospective buyers to other internet services when looking to close a deal, often through hyperlinks to other services. For example, a study which focused on Twitter showed that suspect accounts posted URLs that linked to either online pharmacies supplying controlled substances, or to social media services, blogs, user forums and affiliate marketing (an online advertiser which collects fees for redirecting user traffic to e-commerce platforms), to sell prescription opioids.[1575]

13.65 Petersen et al. found that drug-related posts often featured comments about how to contact sellers outside the service, such as hyperlinks to websites, email addresses or through private messaging services.[1576] QR codes can provide the same functionality as a hyperlink – taking users from one online location to another without explicitly stating what the content of that new location will be in the QR code itself – and have been identified as one route by which websites selling potentially illegal substances have been advertised and users encouraged to visit the sites.[1577]

13.66 Third-party linking services can also often be used to facilitate the sale of illegal substances by moving users from one service to another.[1578]

Recommender systems

Network recommender systems

13.67 Recommender systems have functions beyond suggesting content that a user is likely to find engaging. User-to-user services where users have the option to 'follow' other users or become friends with them, may use recommender systems designed to help users find people they are likely to know (for example, a mutual friend or an old school friend) and likely to want to connect (or reconnect) with. Studies have found that network recommender systems on services can expose the users to individuals offering drugs online and introduce potential buyers to suppliers if they have connected with similar accounts. It is understood that network recommender systems are designed to promote user profiles to users based on, for example, mutual connections or shared interests.

13.68 The Volteface research indicates that once a user connects with a dealer that is openly promoting drugs, they can be exposed to more potential dealers through the 'suggested friends' function which can recommend other dealers. Once researchers connected with

---

[1573] Commission On Combating Synthetic Opioid Trafficking. 2022. Commission on Combating Synthetic Opioid Trafficking: Technical Annexes [accessed 17 October 2022].
[1574] Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, *International Journal of Drug Policy*, 63, 101-110. [accessed 3 June 2023].
[1575] Mackey et al., 2017. Twitter-Based Detection of Illegal Online Sale of Prescription Opioid. [accessed 17 October 2022].
[1576] Petersen et al. 2021. #studydrugs-Persuasive posting on Instagram. [accessed 17 October 2022].
[1577] For example, Metropolitan Police warned Londoners in August 2024 not to scan QR codes promoting illicit cannabis products, with the QR codes taking users to fully operational online marketplaces selling a variety of likely illegal drugs products. Source: Thurston, A. 2024. Londoners warned not to scan mysterious QR codes linking to slick website selling cannabis, MyLondon, 24 August. [accessed 31 October 2024].
[1578] Center for Countering Digital Hate, 2023. TikTok's Toxic Trade. [accessed 25 August 2024].

users who were suspected of drug dealing, they were suggested 'mutual friends' of dealers who were also suspected of dealing. Additionally, researchers saw more user profiles suspected of drug dealing appear in the search bar because of mutual user connections.[1579]

13.69 It is important to acknowledge here that the means of sharing content that promotes drugs matters. If dealers are using messaging functionalities with end-to-end encryption, then it is unlikely to influence the recommender system. When the primary means of encountering the content or users offerings drugs is through functionalities such as these, recommender systems may not play a significant role in suggesting dealers.

Risk factors: Business model

13.70 We are not aware of any evidence on how different revenue models may affect the risks of harm, so we have not assessed which models are relatively high risk or compare relative levels of risk.

[1579] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].
This study found that, "once a few drug dealer accounts had been followed, the platforms would soon start 'suggesting' other drug dealer accounts to follow."

14. Firearms, knives and other weapons

Summary analysis for firearms, knives and other weapons offences: how harm manifests online and risk factors

This chapter summarises the risks of harm to individuals that could happen due to several offences linked to the buying and selling of firearms, knives and other weapons online.

The risks of harm to individuals from these offences are broad. Publication of material in connection with the marketing of knives and the sale or offering to supply firearms, knives and other weapons may result in violent crime, with the most extreme consequence being loss of life.

In the year ending March 2023, 19,555 cautions and convictions were made for possession of a knife or offensive weapon; 18% (nearly 1 in 5) of the cases involved juveniles aged between 10 and 17. While it is difficult to report how many of these knives were originally bought or marketed online, the glamourisation of weapons to young people is of particular concern. Research has identified that more than half of teens have said they had seen real-life acts of violence on social media in the past 12 months and that young people are displaying positive attitudes towards the accessibility and 'coolness' of knives.Offenses in relation to firearms are less prevalent online and appear to operate in closed networks, such as encrypted messaging services. The online sale and subsequent importing of prohibited front-venting blank firing firearms continues to be a focus for

Service type risk factors:

Our evidence points to user-to-user (U2U) services as significant facilitators of these priority offences., However, the role of these service types varies across different offences. For facilitating the supply of illegal knives and sharp weapons our evidence suggests **social media** and **direct messaging services** are particularly important. **Online marketplaces and listing services** and **discussion forums and chat rooms** are prominent service types in the facilitation or commission of weapons offences related to various weapons.

User base group:

The involvement of young people inadvertently and accidentally in knife and weapon crime signifies that **age** is a potential risk factor. Likewise, crime data for knife and weapons offenses shows that **gender** could also be an important influencing feature, with most offenders and victims of weapons offences being men and boys.

Functionalities and recommender systems risk factors:

Unlike the service types the role of functionalities and recommender systems varies less across different offence types. The ability to generate **anonymous user profiles and use direct and encrypted messaging** are key functionalities that facilitate the marketing, sale and purchasing of firearms, knives and other weapons by allowing a direct – privacy preserving – channel of communication between purchaser and seller. **Posting goods or services for sale** particularly supports the ability to sell and purchase knives as there is greater complexity around their legality.

By association with other buying and selling offences (the supply of drugs and psychoactive substances), it may be possible to infer that, **user-generated content searching** and **commenting on content** and **tagging users** may be risk factors associated with the sale, hire, purchase and marketing of firearms, knives and other weapons.

Introduction

14.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the firearms, knives and other weapons offences listed under 'Relevant offences'; and
- the use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

14.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible, consider the effect of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or encountered in combination with content of a different kind.

14.3 The absence of research in theses online offences has meant that we have had to rely heavily on law enforcement expertise to develop our evidence base.

14.4 The firearms, knives and other weapons offences cover various matters relating to the online sale of a range of firearms, knives and other weapons. In the UK, firearms and certain offensive weapons are classified as restricted or prohibited. Prohibited firearms and offensive weapons are subject to the strictest limitations on sale.

14.5 Although we have not considered specific evidence about the dark web for this chapter, we acknowledge it continues to play a role in firearms, knives and other weapons offences.

Relevant offences

14.6 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding firearms, knives and other weapons offences, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

14.7 There are variations in criminality across the firearms, knives and other weapons offences covered within the Act. Some weapons are completely prohibited and therefore it is

completely illegal to possess, sell and supply those weapons. These are often classified as "offensive weapons" and include items such as Butterfly Knives, Zombie Knives and Swords. However, there are other instances, where the Act includes items that can be legal to possess, sell and supply, but only under certain circumstances. For example, some items are commonly prohibited to under 18-year-olds (for example, kitchen knives), whilst others may also require certification to own (for example, a firearm).

14.8 The priority offences relating to firearms, knives and other weapons include the following:

- Possessing, purchasing or acquiring a firearm or ammunition (including air weapons and shotguns) without certificate[1580]
- Dealing in firearms or ammunition (including air weapons and shotguns) by way of trade or business without being registered[1581]
- Sale of firearms or ammunition (including air weapons and shotguns) to a person other than a registered dealer or to a person without certificate[1582]
- Purchase or hire of firearms and ammunition by a person under the age of 18[1583]
- Sale and supply of firearms or ammunition to underage people[1584]
- Sale and supply of firearms or ammunition to persons previously convicted of a crime[1585]
- Purchase of an imitation firearm by a person under the age of 18[1586]
- Sale and supply of imitation firearms to underage persons[1587]
- Sale of realistic imitation firearms[1588]
- Unlawful marketing of knives and publication of material in connection with the unlawful marketing of knives[1589]
- Possessing, purchasing or acquiring, or manufacturing, selling or transferring prohibited weapons[1590]
- Sale or hire of a crossbow to a person under the age of 18[1591]

[1580] Section 1(1) and Section 2(1) of the Firearms Act 1968; section 2 of the Air Weapons and Licensing (Scotland) Act 2015 (asp 10).
[1581] Section 3(1) of the Firearms Act 1968; Article 24 of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3); section 24 of the Air Weapons and Licensing (Scotland) Act 2015 (asp 10).
[1582] Section 3(2) of the Firearms Act 1968; section 24 of the Air Weapons and Licensing (Scotland) Act 2015 (asp 10); Article 37(1) of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3).
[1583] Section 22(1) of the Firearms Act 1968.
[1584] Section 24 of the Firearms Act 1968.
[1585] Section 21(5) Firearms Act 1968; Article 63(8) of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3)).
[1586] Section 24(A) of the Firearms Act 1968; Article 66A of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3).
[1587] Section 24(A) of the Firearms Act 1968; Article 66A of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3).
[1588] Section 36(1)(c) of the Violent Crime Reduction Act 2006 (sale etc of realistic imitation firearms).
[1589] Section 1 and Section 2 of the Knives Act 1997.
[1590] Section 5(1), (1A) or (2A) of the Firearms Act 1968; Article 45(1) or (2) of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3)).
[1591] Section 1 of the Crossbows Act 1987.

- Purchase or hire of a crossbow by a person under the age of 18[1592]
- Manufacture, sale, hire, offer for sale, expose, possess for sale or hire, lend or give to another person banned knives or offensive weapons[1593]
- Sale of knives or other articles with blade or point to underage persons[1594]
- Importation of banned knives, offensive weapons or realistic imitation firearms[1595]

14.9 The Online Safety Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of the offences listed above (and, in relation to offences in Scotland, being involved art and part in the commission of those offences), where applicable.

14.10 Examples of firearms, knives and other weapons offences may include posting weapons for hire or sale on online marketplaces using both text and images. It could also include the sale of firearms, imitation firearms and knives to a person under the age of 18, the unlawful marketing of knives, and publication of material in connection with the marketing of knives.

14.11 Our evidence shows that U2U services can facilitate the sale and purchase of firearms, knives and other weapons by allowing users to post items for sale and facilitate communication between suppliers and buyers. Marketing, in the context of this chapter, does manifest online as advertisements or listings. The marketing may be through text, or images such as explicit animations may be used to indicate that the knife is suitable for combat.[1596]

14.12 For more details on the offences and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance or ICJG.

How firearms, knives and other weapons offences manifest online

14.13 This section is an overview which looks at how firearms, other weapons, and knife offences manifest online, and how individuals may be a risk of harm. While there are similarities in how these buying and selling offences occur, there is a clear distinction in how these illegal weapons are acquired. While knives and other weapons can be widely sold legally in the UK for use as household items, there are greater restrictions against the ownership of firearms.

14.14 To put the risks of harm into context, according to a National Crime Agency (NCA) assessment, few firearms are sold illegally via the clear web.[1597] The clear web tends to be used by people who are not part of a criminal network or who choose to avoid traditional routes to purchase weapons. This may also include an element of physical interaction

[1592] Section 2 of the Crossbows Act 1987.
[1593] Section 1(1) of the Restriction of Offensive Weapons Act 1959; Section 141(1) of the Criminal Justice Act 1988; Article 53 of the Criminal Justice (Northern Ireland) Order 1996 (S.I. 1996/3160 (N.I. 24)).
[1594] Section 141A of the Criminal Justice Act 1988; Article 54 of the Criminal Justice (Northern Ireland) Order 1996 (S.I. 1996/3160 (N.I. 24)).
[1595] Section 1(2) of the Restriction of Offensive Weapons Act 1959; section 141(4) of the Criminal Justice Act 1988; section 36(1)(d) of the Violent Crime Reduction Act 2006 (sale etc of realistic imitation firearms).
[1596] Crown Prosecution Service, 2022. Offensive Weapons, Knife Crime Practical Guidance, Marketing of Knives section. [accessed 18 November 2024].
[1597] Publicly accessible internet. Often referred to as the 'Surface Web'.

between the supplier and the buyer.[1598] Contrastingly, knives and other weapons are more accessible on the clear web because there is a larger variety of legal products. However, regarding the sale of knives and other weapons, investigative reporting from Which?, a consumer protection organisation, highlighted that third-party sellers on large online marketplaces have advertised for sale various prohibited weapons, including various kinds of knives, to UK consumers.[1599]

14.15 Likewise, law enforcement authorities have discovered a variety of prohibited weapons advertised for sale online and delivered to the UK on a large e-commerce website and, online marketplaces.[1600] Items have included irritant sprays, electric shock devices,[1601] front-venting blank-firing firearms.[1602]

14.16 Another discovery by law enforcement authorities highlights the abuse of the legitimate weapons market where bulk orders of non-prohibited knives can be made online from outdoor/sporting goods websites, located via search services, and subsequently sold on online via social media platforms.[1603]

14.17 The Crown Prosecution Service highlights that individuals marketing knives may use slang terms and suggestive messaging and phrasing which advocates the possession of a knife to avoid being a victim of a serious assault.[1604]

Risks of harm to individuals presented by firearms, knives and other weapons offences online

14.18 Supplying or offering to supply firearms, knives and other weapons can result in violent crime, with the most extreme consequence being loss of life.

14.19 While the NCA's assessment highlights that the level of firearm-related crime in the UK is low[1605], addressing the potential for harm from urban street gangs, organised crime groups and potential terrorists remains a priority for UK law enforcement.[1606] [1607]

[1598] National Crime Agency (NCA), 2024. Illegal firearms. [accessed 22 October 2024].

[1599] Which?, 2022. Illegal weapons for sale on AliExpress, Amazon, eBay and Wish, Which? warns. [accessed 22 October 2024].

[1600] NCA response to the November 2023 Illegal Harms Consultation.

[1601] Commonly referred to as a 'Stun Gun' and by the manufacturer brand 'TASER'.

[1602] Self-loading hybrid firearms consisting of approximately 80% 3D-printed components combined with easily accessible metal non-firearms parts. Source: Home Office, 2023. Consultation document (accessible). [accessed 20 September 2023].

[1603] Ofcom / Law Enforcement event / meeting 6 June 2024.

[1604] Crown Prosecution Service (CPS), 2023. Offensive Weapons, Knife Crime Practical Guidance, Marketing of Knives section. [accessed 22 October 2024].

[1605] NCA, 2024. Criminals still want to acquire and use original lethal purpose weapons but they are finding them more difficult to obtain. [accessed 22 October 2024]. For information on 3D Printed Firearms, please see the Terrorism chapter of the Register of Risks.

[1606] There is no overarching consensus on defining a USG. According to the Centre for Social Justice's (CSJ) 2009 report 'Dying to Belong', an Urban Street Gang is defined as *"a relatively durable, predominantly street-based group of young people, who see themselves (and are seen by others) as a discernible group; engage in criminal activity and violence; lay claim over territory (not necessarily geographical but can include an illegal economy territory); have some form of identifying structural feature; and are in conflict with other, similar, gangs".* Source: Centre for Social Justice, 2009. Dying to Belong. [accessed 20 September 2023].

[1607] Organised crime groups are defined as a group of "*members who plan, coordinate and carry out serious crime on a continuing basis. Their motivation is often, but not always, financial gain. Many OCGs are loose networks of criminals who come together for a specific criminal activity, acting in different roles depending on their skills and expertise".* Source: CPS, 2021. Gang related offences – Decision making in. [accessed 20 September 2023].

14.20 The online sale and subsequent importing of prohibited front-venting blank firing firearms continues to be a focus for the NCA due to the potential ease of conversion to live fire and their ability to discharge noxious substances. [1608] Top-venting blank firing firearms, which are legal to purchase and import to the UK, are currently the most common conversions recovered by law enforcement. [1609] This is indicative of the acute risk of harm from such weapons being marketed and sold. [1610]

14.21 Incidents involving other weapons are significantly higher. In the year ending March 2023, there were around 50,500 offences involving a sharp instrument in England and Wales (excluding Devon & Cornwall). In the same year, 19,555 cautions and convictions were made for possession of a knife or offensive weapon; 18% (nearly 1 in 5) of the cases involved juveniles aged between 10 and 17. [1611] It is important to note that it is not possible to establish how many of these knives were originally bought or marketed online.

14.22 The glamourisation of weapons to young people is of particular concern. A study in Scotland found that young people were showing concerningly positive attitudes towards the accessibility and 'coolness' of knives when shown a selection of knife images. It also found that sensationalising images of knives may lead to a climate of fear, increasing paranoia, and potentially inspiring people to carry knives for defence purposes. [1612]

14.23 Children's exposure to and awareness of weapons can be far-reaching, affecting school absenteeism and risking physical and mental harm. In a survey by the Youth Endowment Fund, 55% (more than half) of teens said they had seen real-life acts of violence on social media in the past 12 months. [1613]

Evidence of risk factors on user-to-user services

14.24 We consider that the risk factors listed here are likely to increase the risks of harm relating to firearms, knives, and other weapons offences.

Risk factors: Service types

Messaging services

14.25 Evidence suggests that direct messaging services play a role in the sale and possession of firearms, knives and other weapons. Law enforcement has identified that direct messaging services are often used to facilitate the sale of weapons [1614], and commonly seen social media posts displaying weapons for sale direct users to contact sellers via direct messaging

[1608] NCA response to the November 2023 Illegal Harms Consultation.
[1609] Collectively, front-venting and top-venting blank firers are known as 'convertible blank firers'. Source: NCA, 2024. Firearms threat assessment 2024. [accessed 14 October 2024].
[1610] NCA, 2024. Illegal firearms. [accessed 11 September 2024].
[1611] House of Commons Library, 2023. Knife Crime in England and Wales: Statistics, p.5. [accessed 18 November 2024].
[1612] Cogan, N., Chin-Van Chau, Y., Russell, K., Linden, W., Swinson, N., Eckler, P., Knifton, L., Jordan, V., Williams, D., Coleman, C., and Hunter, S., 2021. Are images of seized knives an effective crime deterrent? A comparative thematic analysis of young people's views within the Scottish context, *PsyArXiv Preprints*. [accessed 18 November 2024].
[30] Youth Endowment Fund, 2022. Children, violence and vulnerability 2022. [accessed 20 September 2023].
[1614] Ofcom / Law Enforcement meeting, 6 June 2024.

to move forward with a purchase.[1615] It has been suggested that these private messaging platforms can be complex and time consuming to investigate, making them popular within criminal networks.[1616] Encrypted messaging software has also been recognised as a methodology for facilitating the sale of weapons in support of international conflicts.[1617]

Social media services

14.26 Our evidence indicates that social media services can be linked to the sale and possession of knives and other weapons. On these services, images and videos are often posted glorifying the use of knives and weapons, as well as showing such weapons for sale.[1618] [1619] It is due to this glamorisation of knife crime on social media accounts that recent youth knife crime fatalities have been linked directly to social media.[1620] In 2021, the Crown Prosecution Service identified how gangs are using social media to glamorise the use of weapons. They explained that the instant nature of social media means that disputes can be easily shared with a large audience, increasing the perceived need of an opposing gang to retaliate.[1621] Researchers have also linked the increased visibility of knife crime on social media to the potential desensitisation of knife crime imagery amongst young people.[1622]

Marketplaces and listing services, discussion forums and chat rooms

14.27 Our evidence points to online marketplaces and listing services as a prominent service type used in the facilitation or commission of these offences. The NCA's National Strategic Assessment of Serious and Organised Crime report, which includes analysis of the impact of firearms, knives and other weapons in the UK, states that online forums, auctions, and online marketplaces are online spaces where the trade of illegal firearms takes place in the UK and many EU countries.[1623]

14.28 An investigation by Which? in 2022 also found that third-party sellers were listing illegal weapons across a variety of popular online marketplaces.[1624]

[1615] "Aman described seeing a range of things for sale, most commonly weapons or drugs, posted by other Snapchat users to their Stories. He said he sees the items with price tags attached to them, often laid out in someone's house. "They just lay it down on a bed or something. A lot of knives or knuckledusters, with the money sign and the amount it is. And then just says, 'Text me if you wanna buy it.' They ask you to text them on Snapchat. After that, they'd get in contact in real life and sell to each other." Source: Revealing Reality, 2023. Anti-social Media. [accessed 11 September 2024].
[1616] Home Office, 2023. Consultation on new knife legislation proposals to tackle the use of machetes and other bladed articles in crime. [accessed 11 September 2024].
[1617] The Economist, 2024. How encrypted messaging apps conquered the world. [accessed 11 September 2024].
[1618] Research by Revealing Reality that explored the content vulnerable children were seeing on Snapchat highlighted that many of the children in the sample had seen content involving firearms, knives and other weapons, including people brandishing these weapons, displaying or using knives or weapons in fights, or advertising weapons for sale. Source: Revealing Reality, 2023. Anti-social Media. [accessed 11 September 2024].
[1619] Sky News, 2024. Teens buying knives illegally online as criminals 'move with digital age'. [accessed 11 September 2024].
[1620] Spring, M, 2022. A social media murder: Olly's story, BBC News, 20 June. [accessed 11 September 2024].
[1621] The Crown Prosecution Service, 2021. Gang related offences - Decision making in. [accessed 11 September 2024].
[1622] Mayor of London, 2023. Study shows impact of knife imagery not universal, but is more profound for some young people affected by violence. [accessed 11 September 2024].
[1623] NCA, 2024. Illegal firearms. [accessed 22 October 2024].
[1624] Which?, 2022. Illegal weapons for sale on AliExpress, Amazon, eBay and Wish, Which? warns. [accessed 22 October 2024].

Risk factors: User base

User base demographics

14.29 The following section outlines primary evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual and complex, involving multiple factors.

14.30 Beyond the evident risk of harm to victims that arises from access to and use of firearms, knives and other weapons in the UK, age of users may be a risk factor for exposure to content that might amount to a weapons-related offence. Sentencing statistics from the Ministry of Justice show that in the year ending March 2023, Juveniles (aged 10 to 17) were the offenders in 18% (nearly 1 in 5) of cases where someone was sentenced for a knife or offensive weapons offence.[1625] At the same time, there is also some indication that children and young people in particular are exposed to online content involving weapons, with 1 in 4 (24%) 13 to 17 year olds in a 2022 survey reporting they had encountered content online that involved "Children or young people carrying, promoting, or using weapons (e.g. a knife, screwdriver or club)".[1626]

14.31 There is also evidence to suggest a significant gender skew in relation this harm, with men and boys consistently more likely to be the perpetrators and the victims of offences concerning firearms, knives and other weapons.[1627] Crime data for the year ending March 2024, shows that over 90% (more than 9 in 10) of convictions or cautions for knife and offensive weapons offences were for men and boys.[1628] We expect crimes facilitated online reflect these trends.

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles

14.32 Evidence suggests that anonymity is strongly favoured in illicit commodity markets and that the perceived risk to the seller from law enforcement authorities has a significant impact on the choice of service they operate from.[1629] [1630] From this, it can be inferred that anonymous user profiles are a risk factor.

Transactions and offers

Posting goods or services for sale

14.33 As mentioned above, an investigation by Which? found that knives and other weapons were being posted for sale on popular online marketplaces. Which? said that it was easily able to find more than one potentially lethal item on each site it looked at, *"at prices*

[1625] MOJ, 2024. Knife and Offensive Weapon Sentencing Statistics: July to September 2023. [accessed 11 September 2024].
[1626] Youth Endowment Fund, 2022. Children, violence and vulnerability 2022., p. 91. [accessed 17 October 2024].
[1627] ONS, 2024. Homicide in England and Wales. [accessed 11 September 2024].
[1628] ONS, 2024. Knife and Offensive Weapon Sentencing Statistics: January to March 2024. [accessed 11 September 2024].
[1629] Bakken, S.A. and Demant, J.J., 2019. Sellers' risk perceptions in public and private social media drug markets. [accessed 22 October 2024].
[1630] Van der Sanden, R., Wilkins, C., Rychert, M. and Barratt, M. J. 2022. 'Choice' of social media platform or encrypted messaging app to buy and sell illegal drugs, *Internal journal of drug policy* 108. [accessed 22 October 2024].

starting from as little as 49p."[1631] Often prohibited items sold online in the UK are made available through international services with users who operate under different legislation.[1632]

Direct messaging

14.34 Research suggests direct messaging functionalities can be used to sell knives and other weapons online.[1633] This is similar to other buying and selling offences, such as those outlined in Drugs and psychoactive substances chapter, where direct messaging can be used to progress a transaction between the buyer and the seller after initiating contact through a variety of services.[1634] The use of direct messaging in the sale of knives to a person under 18 can be time-consuming for law enforcement to investigate, on occasion this can lead to a failure to prosecute.[1635]

Encrypted messaging

14.35 Law enforcement investigations have identified how encrypted messages have been used in the sale and supply of firearms, knives, and other weapons. The end-to-end encryption of the messaging allows users to sell and supply prohibited items without being easily detected.[1636] The international nature of this communication method can also increase the scale of the illegal activity, such as facilitating the illegal sale of firearms in support of international military efforts.[1637] In addition, the security provided by encrypted messaging can often introduce users to networks of criminal activity, increasing exposure to further illegal harms.[1638]

Commenting on content and user tagging

14.36 As detailed in the Drugs and psychoactive substances chapter direct contact between a prospective buyer and seller can be enabled via the 'comments' feature, related to a post. This may also be true for the supply of firearms, knives and other weapons. Through comments, there is also the possible added level of connectivity via the ability to bring the specific comment to the seller's attention by tagging them.[1639]

Content exploring

User-generated content searching

14.37 Given the role that services on which user-generated content is used to offer for sale weapons and other prohibited items play in enabling users to find and potentially purchase weapons, the ability to search specifically for this content is an important functionality.[1640]

[1631] Which?, 2022. Illegal weapons for sale on AliExpress, Amazon, eBay and Wish, Which? warns.

[1632] Petrakos, K, 2024. Warning as machetes, knives and swords for sale online for as low as £1.17, i News, 25 August . [accessed 17 October 2024].

[1633] Revealing Reality, 2023. Anti-social Media. [accessed 11 September 2024].

[1634] Van der Sanden, R., Wilkins, C., Rychert, M. and Barratt, M. J., 2022.

[1635] Home Office, 2023. Consultation on new knife legislation proposals to tackle the use of machetes and other bladed articles in crime. [accessed 11 September 2024].

[1636] Campbell, D. 2022. Encrypted messaging system used to procure gun for murder, court hears. *The Guardian*, 8 February. [accessed 17 October 2024].

[1637] The Economist, 2024. How encrypted messaging apps conquered the world. [accessed 11 September 2024].

[1638] Sherlock, G, 2024. Man who used secret chat to sell guns and drugs jailed, BBC News, 20 August. [accessed 17 October 2024].

[1639] Volteface, 2019. DM for Details: Selling Drugs in the Age of Social Media. [accessed 22 October 2024]

[1640] See 'Social media services', 'Marketplaces and listing services, discussion forums and chat rooms' earlier in this chapter

A Which? investigation found that it was easy to conduct simple searches for banned offensive weapons on popular online marketplaces and that specific characters were often used within the item's title to avoid detection.[1641] Research has also identified that the absence of age verification on certain services has enabled under-18s to purchase weapons prohibited for that age group which they have found and bought through a search for user-generated content in which knives and weapons are posted for sale.[1642] [1643]

Risk factors: Business models and commercial profiles

14.38 No specific evidence was found on how business models may influence risks of harm to individuals for these offences.

[1641] Which?, 2022. Illegal weapons for sale on AliExpress, Amazon, eBay and Wish, Which? warns. [accessed 22 October 2024].
[1642] Busby, M. 2019. Knives being sold via Facebook without any age check. *The Guardian*, 9 August. [accessed 22 October 2024].
[1643] Which?, 2023. Illegal weapons and age-restricted items sold without checks on Temu. [accessed 22 October 2024].

15. Encouraging or assisting suicide (or attempted suicide)

Warning: This chapter contains discussion of suicide, self-harm and eating disorders.[1644]

Summary analysis for encouraging or assisting suicide (or attempted suicide): how harm manifests online, and risk factors

This offence takes place when an individual intentionally encourages or assists a person to (attempt to) end their life. Ofcom's 2024 Online Experiences Tracker found that 4% of UK internet users reported seeing or experiencing content 'promoting suicide' in the past four weeks.[1645] Younger respondents were more likely to see or experience this content, with 6% of 13-to-24-year-olds, 8% of 18-to-24-year-olds and 5% of 25-to-34-year-olds, compared to 3% of those in age groups 35 or older.[1646] The physical and psychological harms that can arise from these offences are severe and can include long-term mental health concerns, eating disorders, physical harm to oneself, and death. Harm from these offences can affect both viewers of the content and the user posting the content themselves.

The role of online content in encouraging or assisting suicide must also be understood in the context of increasing rates of suicide in the UK. The Office for National Statistics (ONS) estimates that the age-standardised suicide rate in England and Wales increased by 15% between 2010 and 2022.[1647]

Content related to suicide is extremely sensitive; while there may be users who post this content to cause harm to others, some users may post this content to find supportive communities, to express their own experiences as part of a healing process or to attempt to help others. Users posting and engaging with this type of content can include those in vulnerable circumstances who are themselves dealing with thoughts of suicide or self-harm, as well as those who have recovered or are recovering from mental health challenges.

There are ethical and legal limitations to conducting research into this type of content, and research has often relied on correlational or qualitative methods for insights into risk factors.

Service type risk factors:

Discussion forum and chat room services can act as spaces where suicide is assisted or encouraged. They may be exploited by individuals with the intent to

[1644] If you need support, please check the following websites: NHS, 2023. Help for suicidal thoughts.; NHS, 2023. Where to get help for self-harm; Beat, 2023. Helplines for eating disorder support.

[1645] This may include content that could be deemed illegal.

[1646] Ofcom, 2024. Online Experiences Tracker 2024. [accessed 18 November 2024]. Fieldwork was carried out in a four-week period in May and June 2024.

[1647] Office for National Statistics, 2024. Suicides in England and Wales: 2023 registrations. [accessed 15 August 2024].

cause harm or distress among users who are experiencing thoughts of suicide. This may be particularly true of services that facilitate discussions on niche or specialised content among smaller groups of users, which could include suicide or self-harm content. However, discussion forums and chat rooms may also be used by individuals experiencing mental health difficulties to connect with other users for support and guidance. **Social media services** can allow potential perpetrators to disseminate suicide related content. Users can view suicide or self-harm content on these types of services, particularly through the creation of user groups on social media services.

Services that allow users to build **online communities** are also a risk factor, as online communities can act as spaces where suicide is promoted or encouraged.

User base risk factors:

Small and large **user base sizes** can pose risks for different reasons. With a larger user base, more people risk encountering this content, while smaller user bases can encourage the sharing of specialised and extreme content relating to suicide.

Users who are in vulnerable circumstances such as such as those suffering with their **mental health** and who might be experiencing thoughts of suicide or self-harm are more likely than other users to be at risk from the effects of this type of content.

Age is also a potential risk factor. Our evidence suggests that **young people** are more likely to encounter this content, to use the internet for suicide-related purposes, and to be susceptible to copycat behaviour.

Functionalities and recommender systems risk factors:

Commenting on content is a risk factor, there is evidence of people using comments to encourage the suicide of the person that distributed the content. Commenting on content intersects with other risk factors such as **livestreaming** and **posting content** to create high-risk context. For example, livestreaming is a risk factor that has been used to share real-time acts of suicide or self-harm. While livestreaming these activities is not in itself illegal, the social functionality attached to the livestream, including commenting on the livestream or in **user groups** connected to the livestream can be used to encourage the suicide or self-harm depicted on the livestream. Similarly, comments on posts related to suicide can encourage the user, who may be experiencing thoughts of suicide, to attempt to take their own life.

Anonymous user profiles appear to be a risk factor. Some users may feel more confident in sharing content depicting or discussing harmful themes if they cannot be identified. However, users may also feel that anonymity allows them to talk more openly about their own thoughts of suicide, and to connect with others with similar experiences. Therefore, anonymity can both pose risks and confer potential benefits to those seeking help.

The **ability to post content** and **re-post or forward content** can enable and benefit users to connect with others who are experiencing similar thoughts or behaviours, but it can also be used to disseminate harmful suicide and self-harm content.

Content recommender systems can also be a risk factor. The way in which recommender systems are designed can influence the extent to which harmful (and potentially illegal) content is recommended to users. Research suggests that where there are vulnerable users who are engaging with harmful content, such as self-harm or suicide content, recommender systems are more likely to create a 'filter bubble' or 'rabbit hole.' This may lead to users discovering more content that is harmful or distressing, as well as potentially illegal. If a user is primarily engaging with harmful content, then this is likely to create a filter bubble where the user is recommended more harmful content, while other content is deprioritised.

Other functionalities risk propagating this offence. **Content tagging** such as hashtags can help evade content moderation techniques on suicide or self-harm content, because groups of users create hashtags that differ from those that may be blocked as harmful. **Group messaging** can also enable users to contact one another and can encourage harmful behaviour in a group setting.

Introduction

15.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to the offence of encouraging or assisting suicide detailed under 'Relevant offences'; and
- The use of these services for the commission and/or facilitation of this offence (collectively, the 'risks of harm').

15.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible, consider the effect of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.[1648]

15.3 Suicide is not a criminal offence; nor is discussing or portraying suicide in any way which does not amount to encouraging or assisting suicide. An individual attempting to do this will not and should not be penalised. Individuals suffering with their mental health or in otherwise vulnerable circumstances, including those dealing with thoughts of suicide and self-harm, should be able to seek support without fear of negative consequences.

15.4 Services should be aware that content relating to suicide varies and is not always shared with intent to encourage or assist suicide. Users who share suicide or self-harm content may themselves be vulnerable, using online spaces to express their feelings and seek

[1648] As with other chapters, we have considered evidence of suicide content from a variety of sources, including information provided by services, academic literature, third-party research and civil society in general. Some of this evidence relates to content which may not necessarily mirror, or is broader than, the criminal definitions of these offences.

support by connecting with others who may be having similar experiences. Service providers should therefore be mindful of this distinction when assessing this type of content, considering the risks of harm to the user who shares the content as well as to other users.

15.5 In this chapter we explore the evidence related to an increased risk of encountering content or activity online that may amount to the offence of encouraging or assisting suicide. For the purposes of our assessment, including to assess the effect characteristics have on the risks of harm, we treat some content as potentially amounting to illegal content, recognising that whether it is illegal content depends on the intentions of the person sharing the content (see 'Relevant offences' and the Illegal Content Judgements Guidance or ICJG.

15.6 We will occasionally refer to evidence that references self-harm as well as suicide. This is because it is often difficult to draw a clear distinction between the two types of content, and much of the research in this area that contributes to our understanding does not focus solely on one or another type of content.[1649]

15.7 Where data from Ofcom's Online Experience Tracker is included, this is based on participants' self-reported experience of having seen or experienced 'content relating to self-harm or suicide' or 'content promoting suicide', which may not necessarily include content deemed to meet the illegal threshold.

Relevant offences

15.8 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. Regarding suicide, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

15.9 In this chapter, we consider the priority offence of encouraging or assisting suicide.[1650]

15.10 An offence can take place when a person encourages or assists the suicide, or an attempted suicide, of another person. Any content online that intentionally encourages or assists a person to end their life may constitute illegal content.[1651]

15.11 For this offence, it is not necessary for the encouragement or assistance to be targeted towards a specific person or persons. The content also does not need to result in suicide or attempted suicide for it to amount to illegal content.

15.12 For more details on the offences and how service providers can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance or ICJG.[1652]

[1649] Brennan, C., Saraiva, S., Mitchell, E, Melia, R., Campbell, L., King, N. and House, A., 2022. Self-harm and suicidal content online, harmful or helpful? A systematic review of the recent evidence, *Journal of Public Mental Health,* 21 (1). [accessed 10 July 2023].

[1650] Section 2 of the Suicide Act 1961 and section 13 of the Criminal Justice Act (Northern Ireland) 1966 (c. 20 (N.I.)). (c. 20 (N.I.)).

[1651] Samaritans, 2020. Understanding self-harm and suicide content online. [accessed 24 May 2023].

[1652] While some of the content referred to in this chapter may cause harm/distress, it may not necessarily meet the criminal threshold. Please refer to the Illegal Content Judgements Guidance to assess whether content amounts to illegal content.

How encouraging or assisting suicide manifests online

15.13 This section is an overview which looks at how the offence of encouraging or assisting suicide manifests online, and how users may be at risk of harm.

15.14 To put the risks of harm from this offence into context, wave 6 of Ofcom's 2024 Online Experiences Tracker found that 4% of UK internet users reported seeing or experiencing content 'promoting suicide' in the past four weeks.[1653] Younger respondents were more likely to see or experience this content, with 6% of 13-to-24-year-olds, 8% of 18-to-24-year-olds and 5% of 25-to-34-year-olds, compared to 3% of those in age groups 35 or older.[1654]

15.15 Two-thirds (66%) of UK adults say they are concerned about the accessibility of 'harmful suicide or self-harm content' online.[1655] This is likely to include content that could be considered illegal. However, due to ethical and practical limitations, it is challenging to identify whether exposure to suicide and self-harm related content online causes increased risk of suicide or self-harm related outcomes.[1656]

15.16 The role of online content in encouraging or assisting suicide must also be understood in the context of increasing rates of suicide in the UK. The Office for National Statistics (ONS) estimates that the age-standardised suicide rate in England and Wales increased by 15% between 2010 and 2022[1657], although we do see different trends in Scotland and Northern Ireland.[1658]

Risks of harm to individuals presented by the offence of encouraging or assisting suicide online

15.17 Suicide and self-harm content can manifest online in various forms, with a range of effects on individuals. Samaritans, a charity that works with people struggling to cope and people at risk of suicide, notes examples of suicide or self-harm content that may pose a risk to individuals.[1659] These include detailed and instructive information about suicide or serious self-harm methods, posts encouraging, glamourising or celebrating suicide or serious self-harm, and graphic images relating to serious self-harm or suicide.[1660] Not all of this content

---

[1653] This may include content that could be deemed illegal.
[1654] Ofcom, 2024. Online Experiences Tracker 2024. [accessed 18 November 2024]. Fieldwork was carried out in a four-week period in May and June 2024.
[1655] This concern is not limited to the impact of this content on children, but also on adults. Four in five (83%) UK adults agree that "harmful suicide or self-harm content" can have a damaging effect on adults as well as children. Source: Samaritans, 2023. Government is failing the public with online safety bill, says Samaritans. [accessed 19 January 2023].
[1656] Arendt, F., Scherr, S. and Romer, D., 2019. Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults, *New Media & Society,* 21 (11-12). [accessed 10 July 2023].
[1657] Office for National Statistics, 2024. Suicides in England and Wales: 2023 registrations. [accessed 15 August 2024].
[1658] Conversely, in Northern Ireland, the age-standardised rate of probable suicides reduced by 36% between 2010 and 2022 - Northern Ireland Statistics and Research Agency, 2023. Suicide statistics 2022 [accessed 12 October 2024]; The age-standardised rate of probable suicides in Scotland in 2022 was 14 per 100,000 people, which was not significantly different from the rate in 2010 (14.7), nor was there any clear positive or negative trend over the years within that range - National Records of Scotland, 2024. Probable Suicides [accessed 12 October 2024].
[1659] Samaritans, 2022. Towards a suicide-safer internet. [accessed 24 May 2023].
[1660] The full list of examples includes: detailed and instructive information about suicide or self-harm methods; posts encouraging, glamourising or celebrating suicide or self-harm; posts from people seeking or encouraging suicide or self-harm pacts; posts relating to suicide or self-harm challenges; graphic images relating to self-harm or suicide; and livestreams or recorded videos of suicidal or self-harming behaviour.

will be illegal, although all has the potential to be harmful. For information on what could be considered illegal, please refer to the ICJG guidance.

15.18 There are at least two distinct groups of users who are likely to be at risk: those who encounter this content unintentionally (for example, when searching for content that overlaps with hashtags used to share harmful suicide-related content, or when such content is served up by a recommender system), and those who may be experiencing thoughts of suicide or serious self-harm and are seeking this type of content. Other users at risk may include those who are looking to disengage from these kinds of content but encounter it again due to their previous online engagement.

15.19 Several studies have further explored the impact of exposure to suicide and self-harm related content. A study of 18 to 29-year-olds found that those who viewed content depicting self-harm on a social media service, either intentionally or by accident, are at a higher risk of suicide and self-harm.[1661] The potential negative effects of this content were also evident in a national survey by Swansea University and Samaritans (where 87% of the sample reported having self-harmed before).[1662] It asked respondents about the impact of seeing or sharing suicide or self-harm content online: one in three (35%) reported a worsening of their mood, with only 2% reporting that this type of content improved their mood.[1663] However, more than half the respondents reported that the impact this content had on them depended on their mood at the time, so the proportion whose mood was negatively affected is potentially higher than 35%.

15.20 It is likely that repeated exposure to suicide-related content within online communities works to normalise the act of suicide[1664], increasing the risk to those users of undertaking an act of suicide.[1665] One reason that the frequency of dissemination and access to this kind of content online is a concern is because suicide or self-harm content may have a 'contagion' effect, whereby being exposed to the idea of suicide increases the risk that those exposed undertake an act of suicide. For example, the Royal College of Psychiatrists identify "the well-known 'contagion' effects of self-harm in inpatient units".[1666] This could also apply to online contexts.

15.21 There is strong evidence that the contagion effect extends to media portrayals of suicide, including fictional suicides. A review of research articles testing the 'Werther Effect' – the hypothesis that popular media portrayals of suicide increase the suicide rate of the audience by imitation – found that about 70% research articles reviewed (69/98) found

---

[1661] The study found that exposure to these depictions of self-harm on Instagram resulted in an 'emotional disturbance' in some users, with this exposure positively associated with psychometric predictors of "(possibly harmful) self-harm and suicidality-related outcomes". Source: Arendt, F., Scherr, S. and Romer, D., 2019.

[1662] The sample included 5,294 individuals aged 16-84 years. Many of the participants in the study were females aged under 25, and so does not represent any population as a whole. Source: Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. [accessed 10 July 2023].

[1663] Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. [accessed 10 July 2023].

[1664] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023. Research Review: Viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms. [accessed 10 October 2024].

[1665] Oexle, N., Valacchi, D., Grubel, P., Becker, T., Rusch, N. 2022. Two sides of the same coin? The association between suicide stigma and suicide normalisation. *Epidemiology and Psychiatric Sciences*, Vol 31. [accessed 10 October 2024]

[1666] Royal College of Psychiatrists, 2020. Technology use and the mental health of children and young people. [accessed 10 July 2023].

evidence for the effect.[1667] Fourteen of the 69 studies confirming the effect found that fictional dramatizations of suicide were linked to, and likely caused, subsequent increases in suicide rates. Most studies also reported that people who experience emotional or mental health challenges, or who share the portrayed suicide victim's demographic characteristics (age, gender, nationality), or otherwise share similar life situations, are more likely to imitate depictions of suicidal behaviour. Additionally, young people and adolescents were found to be the most vulnerable to imitative influences.[1668]

15.22 According to Samaritans, evidence suggests that content presenting suicide behaviours (such as viral suicide and serious self-harm 'challenges', encouraging users to engage in harmful behaviour), may encourage or assist other users to undertake acts of suicide. They state that the contagion effect may become more likely, increasing the risk of imitation, when the viewer overly identifies with the original uploader of the content (for example, if they are at increased risk of thoughts of suicide or serious self-harm).[1669]

15.23 A users' mental state at the time of viewing content detailing suicide methods may also determine the effect that certain types of suicide or self-harm content has on them. NatCen, a UK-based research agency, found that individuals seeing information on how to take one's life had risked exacerbating their suicidal thoughts at a time when the men interviewed in the study were feeling distressed, isolated and confused.[1670]

Evidence of risk factors on user-to-user services

15.24 We consider that the risk factors below are likely to increase the risks of harm relating to encouraging or assisting suicide.

Risk factors: Service types

15.25 Research indicates that the following types of services can be used to commit or facilitate the offence of encouraging or assisting suicide: discussion forums and chat rooms, information-sharing services, social media services, video-sharing services and services that more generally enable community building.

Discussion forums and chat rooms, information sharing services

15.26 Our evidence shows that discussion forums and chat room services can act as spaces where suicide is assisted or encouraged. Although these services can have positive benefits, they can also facilitate discussion and ideation relating to suicide and self-harm, which can escalate into encouragement of suicidal behaviours, including sharing content that can be

[1667] Domaradzki, J. 2021. The Werther Effect, the Papageno Effect or No Effect? A Literature Review. *International Journal of Environmental Research and Public Health*, 18(5). [accessed 10 October 2024]

[1668] Media portrayals also increase the rate of suicides by the same suicide method as that portrayed. Domaradzki, J. 2021. The Werther Effect, the Papageno Effect or No Effect? A Literature Review. *International Journal of Environmental Research and Public Health*, 18(5). [accessed 10 October 2024].

[1669] Samaritans, 2022. Towards a suicide-safer internet. [accessed 24 May 2023].

[1670] NatCen (McManus, S., Lubian, K., Bennett, C., Turley, C., Porter, L., Gill, V., Gunnell, D. and Weich, S.), 2019. Suicide and self-harm in Britain – researching risk and resilience. [accessed 10 July 2023].

harmful or distressing to users.[1671] Users may also specifically post to prompt other users to provide them with information detailing suicide methods.

15.27 A number of deaths in the UK in recent years have reportedly involved chatrooms and forums,[1672] and the Royal College of Psychiatrists has emphasised the normalisation of sharing graphic images of self-harm on discussion forums as a significant concern among those who are already vulnerable.[1673]

15.28 A small-scale qualitative study in the UK looking at 18 to 24-year-olds who previously had suicidal thoughts found that distressing content on online chat groups, blogs, and forums had exacerbated their suicidal feelings, and for some respondents, content on online forums emerged as the main factor in generating negative effects. Although a number of respondents reported that these spaces also enabled users to find help and support.[1674]

15.29 A US study, looking at young people aged 14 to 24 who knew individuals who had attempted, or died by, suicide, and who had themselves experienced hopelessness and suicidal ideation, found that "discussion forums appear to be particularly associated with increases in suicidal ideation."[1675] The respondents cited social media services as key sources of information about suicide content, but the respondents did not associate the services with increases in suicide ideation. However, the research notes that online discussion forums were cited as sources of information and linked with increases in ideation.

15.30 Research indicates that "chatrooms and discussion forums may also pose a risk for vulnerable people by raising the option and then influencing decisions to die by suicide."[1676] Some individuals have reported being encouraged to die by suicide on such services. Researchers suggest that these kinds of conversations can facilitate suicide 'pacts' (where individuals arrange their collective death), create peer pressure to take one's own life, and encourage suicide ideation.[1677]

15.31 Notwithstanding the fact that they sometimes play a role in exacerbating suicidal thoughts, there is also evidence that people sometimes use chat rooms in a way that helps them cope with such thoughts. A UK-based qualitative study with participants who had either previously used the internet for suicide-related purposes, or had been admitted to hospital

---

[1671] There are services such as some discussion forums that are dedicated to suicide or self-harm content. However, a recent inquest has revealed that this content also exists across services that actively prohibit suicide or self-harm content. Source: The Coroner's Service, 2022. Prevention of Future Deaths. [accessed 28 October 2022].

[1672] In the UK between 2001 and 2008, there were at least 17 deaths involving chatrooms or sites that provide advice on suicide methods. Source: Cohen-Almagor, R, and Lehman-Wilzig, S., 2022. Digital Promotion of Suicide: A Platform-Level Ethical Analysis, *Journal of Media Ethics,* 32 (2). [accessed 10 July 2023]. One forum in particular was linked to 50 deaths in the UK, with coroners and police investigations highlighting the role it played and issuing warnings about the site. BBC, 2023. 'Failure to act' on suicide website linked to 50 UK deaths. [accessed 26 September 2024].

[1673] Royal College of Psychiatrists, 2020. Technology use and the mental health of children and young people. [accessed 10 July 2023]. For more information, see the section, 'Risks of harm to individuals presented by the offence of encouraging or assisting suicide online'.

[1674] Bell, J., Mok, K., Gardiner, E. & Pirkis, J., 2017. Suicide-related internet use among suicidal young people in the UK: Characteristics of users, effects of use, and barriers to offline help-seeking, *Archives of suicide research: official journal of the International Academy for Suicide Research*, 22 (4). [accessed 10 July 2023].

[1675] Dunlop, S M., More, E. and Romer, D., 2011. Where do youth learn about suicides on the internet, and what influence does this have on suicidal ideation? [accessed 5 July 2023].

[1676] Cohen-Almagor, R, and Lehman-Wilzig, S., 2022. Digital Promotion of Suicide: A Platform-Level Ethical Analysis, p.112, *Journal of Media Ethics,* 32 (2). [accessed 14 October 2024].

[1677] Cohen-Almagor, R, and Lehman-Wilzig, S., 2022.

following serious self-harm, found that a number of young adults in the sample used discussion forums and chatrooms, information-sharing services (in this case, Q&A websites), and social media services to express their feelings, manage loneliness or engage in dialogue with others.[1678]

Social media services

15.32 The available evidence suggests that social media services can play a role in the dissemination of harmful suicide and self-harm related content. Research shows that users who view self-harm on social media services,[1679] either intentionally or by accident, are at a higher risk of self-harm or suicide.[1680] Dedicated self-harm or suicide groups are also occasionally set up by users on social media services, offering users a chance to discuss topics with other users.[1681] The research also indicates that the ability to post and share content, for example by posting hyperlinks, can escalate content related to suicide and lower the mood of users, particularly on social media services and information-sharing services such as Q&A websites.[1682]

Video-sharing services

15.33 Video-sharing services can also play a role in disseminating suicide and self-harm content. There have been several cases in which livestreaming, a functionality common to video-sharing services, has been used to show users self-harming or ending their life in real time.[1683] Comment threads on video-sharing services, in particular, have been shown to contain content that could amount to assisting or encouraging suicide (see Commenting on content for more information).

Risk factors: User base

User base size

15.34 Services with both large and small user bases pose risks in relation to suicide and self-harm content, for different reasons.

15.35 On the one hand, the larger a service's user base, the greater the number of people who are likely to encounter content on it, particularly where it is amplified through recommender systems, meaning that content can receive substantial amounts of engagement.[1684] (See Commenting on content for more information). This in turn heightens

---

[1678] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[1679] In particular, Instagram, as the focus of this study at the time.

[1680] The study found that exposure to this content resulted in an 'emotional disturbance' in some users, with this exposure statistically related to '(possibly harmful) self-harm and suicidality-related outcomes'. Source: Arendt, F., Scherr, S. and Romer, D., 2019. Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults, p.3, *New Media & Society,* 21 (11-12). [accessed 10 July 2023].

[1681] Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., Purdy, N., Daine, K. and John, A., 2017. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown, *PLoS ONE,* 12 (8). [accessed 10 July 2023]. For more information, see the section, 'Risk factors: functionalities and recommender systems'.

[1682] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[1683] For more information, see 'livestreaming' in the section, 'Risk factors: functionalities and recommender systems'.

[1684] Ekō, 2023. Suicide, Incels, and Drugs: How TikTok's deadly algorithm harms kids. [accessed 11 July 2023].

the risk of contagion effects, as described earlier. (See Risk factors: functionalities and recommender systems for further information)

15.36 Meanwhile, services with a small user base may be more likely to encourage the sharing of more niche or specialised content, which could include suicide or self-harm content.

15.37 Small sites dedicated to the discussion of subject matters related to suicide can present a particularly high risk. One small forum, specialising in providing information about suicide methods, has been linked by coroners reports to 50 deaths in the UK.[1685] Content analysis of the forum showed that 30% of the discussions were about suicide methods and that the site has been linked to increased incidents of previously infrequent methods of suicide.[1686]

User base demographics

15.38 The following section outlines the key evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

15.39 Data suggests that user base characteristics including age, mental health, **ethnicity, sexual orientation,** and **gender** could all play a role in increasing the chance that an individual is exposed to content that encourages suicide online, and potentially take action as a result of that exposure.

15.40 The data referenced here from Ofcom's Online Experience Tracker is based on participants' self-reported experience of having seen or experienced 'content relating to self-harm or suicide' or 'promoting suicide', which may not necessarily include content deemed to meet the illegal threshold.

15.41 Data suggests that the age of users influences their likelihood of encountering suicide-related content. Younger respondents were more likely to see or experience this content, with 6% of 13-to-24-year-olds, 8% of 18-to-24-year-olds and 5% of 25-to-34-year-olds, compared to 3% of those in age groups 35 or older.[1687] The evidence also suggests that younger adults are more likely to experience the contagion effect ('copycat' behaviour)[1688] and to have used the internet for suicide-related purposes (among those who had been in contact with mental health services).[1689]

15.42 A study by Samaritans and Swansea University found that people with a history of self-harm were more likely to report that they were 10 years old or younger when they first viewed self-harm or suicide content online, whereas those with no history of self-harm were more

---

[1685] Cooper, G., Lewis, C., 2023. 'Failure to act' on suicide website linked to 50 UK deaths, BBC, 24 October. [accessed 28 July 2024].
[1686] Sartori, E. 2022. Analyzing Sanctioned Suicide: a case study on pro-choice sites. Università di Padova. [accessed 9 September 2024].
[1687] Ofcom, 2024. Online Experiences Tracker 2024. [accessed 22 November 2024].
[1688] Domaradzki, J. 2021. The Werther Effect, the Papageno Effect or No Effect? A Literature Review. *International Journal of Environmental Research and Public Health*, 18(5). [accessed 10 October 2024].
[1689] In 2011-2018, a national confidential inquiry into suicide and homicide by people with mental illness found that 15% of under-25s (aged 10+) had reported using the internet for suicide-related purposes (e.g. visiting pro-suicide websites) during this time, which was significantly higher than for patients aged 25+ (7%).
Source: University of Manchester, 2021. The National Confidential Inquiry into Suicide and Safety in Mental Health. [accessed 11 July 2023].

likely to report being aged 25+ at the time of first encountering this content.[1690] In response to an Ofcom call for evidence, Samaritans suggested that this may indicate a potential correlation between viewing harmful content at a young age and future harmful behaviour.[1691]

15.43 Children presenting to hospital following self-harm are more likely to have used the internet for purposes related to suicide, including to obtain information on suicide methods, visit pro-suicide websites, communicate suicidality online, and other suicide related purposes. A study conducted between 2013 and 2015 examined the cases of 1,198 individuals who presented following self-harm at one of two Bristol hospitals and provided information about their internet use.[1692] The findings revealed that suicide related internet use[1693] was significantly more prevalent among children (26%) compared to adults (8.4%).[1694] Similarly, a study of UK mental-health patients who died by suicide between 2011 and 2021 found that suicide related internet use was about 2.7 times more likely for people aged 25 years-old or younger compared to all other ages.[1695] A study reviewing coroner inquest data on suicides by young people (ages 10 to 19) between 2014 and 2016 found that suicide related online experience was an antecedent in 24% of those deaths.[1696]

15.44 Although rates of suicide are consistently highest for those aged 40 to 54 years old, rates of suicide among young people have increased the most since 2010[1697]. Per 100,000 10 to 24-year-olds, 5.2 died by suicide in 2023 (561 suicides), a 27% increase over 2010, when there were 4.1 suicides in every 100,000 10 to 24-year-olds (432 suicides). Of particular concern is the sharp upward trend in the number of children aged 10 to 14 dying by suicide in England and Wales.[1698] Eighteen 10 to 14-year-olds died by suicide in 2022, following a steady upward trend over the years since 2010 when just two died. Furthermore, over the same time-period, the suicide rate among 15-to-19-year-olds increased by 65%.[1699]

15.45 Not all of those experiencing suicidal ideation have a history of mental health challenges and may instead be experiencing adverse circumstances. However, neurological and psychological factors are also associated with elevated risk of harm. Ofcom's Online Experiences Tracker (OET) data suggests that internet users with mental health conditions

---

[1690] Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. [accessed 10 July 2023].
[1691] Samaritans response to 2023 Ofcom Call for Evidence: Second phase of online safety regulation: Protection of Children
[1692] Data from Bristol Royal Infirmary (BRI) and Bristol Royal Hospital for Children (BRHC) was analysed. Suicide and Self-Harm Related Internet Use.
[1693] In this study, suicide related internet use (SRIU) is composed of the following recorded activities, percentages are of those that engaged in any SRIU activity: obtaining information on suicide methods (68.9%), visited pro-suicide websites (32.9%), communicated suicidality online (15.9%), other SRIU including purchasing means of suicide online (9.2%). Percentages do not sum to 100% because many individuals engaged in more than one activity.
[1694] Padmanathan, P., Biddle, L., Carroll, R., Derges, J., Potokar, J., & Gunnell, D. (2018). Suicide and Self-Harm Related Internet Use. *Crisis*, 39(6), 469–478. [accessed 10 October 2024].
[1695] Bojanić, L., Turnbull, P., Ibrahim, S., Flynn, S., Kapur, N., Appleby, L., Hunt, I. M. 2024. Suicide-related internet use among mental health patients who died by suicide in the UK: a national clinical survey with case-control analysis. *The Lancet Regional Health – Europe.* Volume 44, September. [accessed 10 October 2024].
[1696] Rodway, C., Tham, S. G., Richards, N., Ibrahim, S., Turnbull, P., Kapur, N., Appleby, L. 2023. Online harms? Suicide-related online experience: a UK-wide case series study of young people who die by suicide. *Psychol Med*. 2023 Jul;53(10):4434-4445. [accessed 10 October 2024].
[1697] Office for National Statistics, Suicides in England and Wales: 2023 registrations. [accessed 11 August 2024].
[1698] Commensurate figures for Scotland and Northern Ireland were not available.
[1699] From 3.1/100,000 to 5.1/100,000. Office for National Statistics, Suicides in England and Wales: 2023 registrations. [accessed 11 August 2024].

are significantly more likely to experience content promoting suicide (7% versus 2% with no diagnosed mental health conditions).[1700] A UK study comparing clinical characteristics of suicide victims between 2011 and 2021 found that, controlling for age and gender, the likelihood of suicide related internet use was higher among those diagnosed with affective, anxiety, and autism spectrum disorders, and those receiving any psychological treatment.[1701] A higher proportion (21%) of autistic patients had used the internet for suicide-related purposes compared to just 7% of non-autistic patients.[1702]

15.46 Other evidence also suggests that those with existing mental health challenges may be more likely to encounter suicide or self-harm content.[1703]

15.47 Additionally, the **ethnicity**, **sexual orientation** and **gender** of users may play a role in increasing the risk of harm related to suicide or self-harm content.

15.48 Ofcom's OET suggests that content promoting suicide is more likely to be experienced by internet users who are of mixed ethnicity (6%) or Black (7%) compared to those who are white (3%).[1704]

15.49 OET data also suggests that this online harm is experienced more by adult internet users who identify as bisexual (8%) versus heterosexual (3%). Furthermore, those who identified as non-binary were significantly more likely to report experiencing content promoting suicide (24%) than males (4%) and females (3%).[1705] A study of coroner inquest data on suicides by young people (aged 10 to 19) between 2014 and 2016 found that those identifying as LGBT were 2.3 times more likely to have suicide related online experience [1706] reported as an antecedent to their suicide than non-LGBT individuals.[1707]

15.50 About 1 in every 4 suicides are women, meaning that men are three times more likely to die by suicide than women.[1708] Nonetheless, the suicide rate of young women and girls is increasing more steeply than that of young men and boys. From 2010 to 2022, there was an estimated increase in suicide rate of 18% (6.1 to 7.2 per 100,000) among males aged 10 to 24, compared to females among whom there was an estimated increase of 48% (2.1 to 3.1

[1700] Ofcom, 2024. Online Experiences Tracker 2024. [accessed 22 November 2024]. Ofcom's OET refers to mental health as 'Anxiety, depression or trauma-related conditions, for example.'

[1701] Lana B., Pauline T., Saied I., Sandra F., Navneet K., Louis A., Isabelle M. H., 2024, Suicide-related internet use among mental health patients who died by suicide in the UK: a national clinical survey with case-control analysis. [accessed 21 June 2024]

[1702] The National Confidential Inquiry into Suicide and Safety in Mental Health, 2024. Annual Report: UK patient and general population data, 2011-2021. [accessed 10 October 2024]

[1703] Ofcom research into how people are harmed online included a case study on a male aged 26-30. He had struggled with his mental health over the lockdown period and as a result had sought information on suicide methods via a search engine, until he reached forums that discussed suicide methods. His poor mental health increased the likelihood that he would experience harm from the content. Ofcom, 2022. How people are harmed online: Testing a model from a user perspective. [accessed 11 July 2023].

[1704] Ofcom, 2024.

[1705] Ofcom, 2024. Online Experiences Tracker 2024. [accessed 22 November 2024].

[1706] Suicide related online experience includes: searching the internet for information on suicide method, visiting websites that may have encouraged suicide behaviour, communicating suicidal ideation or intent online and being bullied online. Rodway, C., Tham, S. G., Richards, N., Ibrahim, S., Turnbull, P., Kapur, N., Appleby, L. 2023. Online harms? Suicide-related online experience: a UK-wide case series study of young people who die by suicide. *Psychol Med*. 2023 Jul;53(10):4434-4445. [accessed 10 October 2024].

[1707] Female *versus male* (OR 1.87, 95% CI 1.23-2.85, *p* = 0.003), LGBT *versus non-LGBT* (OR 2.35, 95% CI 1.10-5.05, p = 0.028).

[1708] Averaging over age-standardised suicide rates for the years between 2010 and 2022. Office for National Statistics, Suicides in England and Wales: 2023 registrations. [accessed 11 August 2024].

per 100,000). A study reviewing coroner inquest data on suicides by young people (ages 10 to 19) between 2014 and 2016 found that females were 1.9 times more likely to have a suicide-related online experience[1709] recorded prior to death.

Risk factors: Functionalities and recommender systems

User identification

Anonymous profiles

15.51 While anonymity has important benefits,[1710] it can also result in users feeling comfortable in sharing or engaging with more harmful or explicit content, thereby increasing the risk of potential illegal content being shared. In other cases, anonymity can embolden 'suicide baiters' – those who wish to encourage the suicide of individuals expressing suicidal ideation – to commit the offence online. [1711]

15.52 Anonymity can result in some users feeling more comfortable sharing explicit content than on their identifiable profiles. A study from the Netherlands found that several adolescents who had ended their lives had created secondary social media accounts under false names, and at least five girls had used these accounts to enter communities anonymously and "to share explicit suicide-related communications." Respondents in the study said this was due to the younger girls being cautious of the potential judgement and consequences from their family or friends if they encountered this suicide-related content.[1712]

15.53 A user's posts on an anonymous user profile can sometimes become more explicit as interest in the profile and associated content grows. One participant in a study by Biddle *et al.* (2018), noted "I created an anonymous Instagram page. At first it was captions and quotes and stuff that I'd find, and I thought were quite good, then once I saw how many people were looking at the page, I started posting pictures of [self-harm] and getting more and more followers and it became addictive. It eventually got shut down, it became pro-self-harm."[1713]

Fake user profiles

15.54 Being able to create fake user profiles can increase the risk of pro-suicide content being disseminated on a service. There are case examples where perpetrators have created false identities, as opposed to simply remaining anonymous, to maliciously encourage others to take their own lives. False identities can be used to create personas that users would be more likely to relate to and be influenced by. For example, a case in the USA involved a potential perpetrator (described as a 'serial predator') who was convicted for using fake

---

[1709] Suicide-related online experience includes: searching the internet for information on suicide method, visiting websites that may have encouraged suicide behaviour, communicating suicidal ideation or intent online and being bullied online. Rodway, C. et al. 2023.

[1710] Anonymity can have benefits in helping some individuals feel more able to express themselves online, particularly users who may be experiencing thoughts of suicide or self-harm.

[1711] Phillips, J G., Diesfeld, K. and Mann, L., 2019. Instances of online suicide, the law and potential solutions, *Psychiatry, Psychology and Law,* 26 (3). [accessed 27 January 2023].

[1712] Balt, E., Mérelle, S., Robinson, J., Popma, A., Creemers, D., Brand, IVD., Bergen, DV., Rasing, S., Mulder, W. and Gilissen, R., 2023. Social media use of adolescents who died by suicide: lessons from a psychological autopsy study, *Child and Adolescent Psychiatry and Mental Health,* 17 (48). [accessed 11 July 2023].

[1713] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.12, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

identities to encourage individuals to join bogus suicide pacts and to assist with suicides by suggesting methods for victims and survivors to use.[1714]

User networking

User groups

15.55 User groups that are dedicated to discussing suicide topics with other users can be created, particularly on social media services. These can be a source of support, but evidence suggests that they can also contain content which glorifies or normalises suicide. User groups within social media services are often not moderated in the same way as support forums, where rules about inappropriate content are more likely to exist.[1715]

User communication

Group messaging

15.56 Group messaging allows users to contact one another and potentially encourage harmful behaviour in a group setting. While our evidence often cites 'chatrooms', because chatrooms are centred around enabling users to message one another as groups, we have used research on chatrooms to draw conclusions surrounding group messaging.

15.57 In the UK between 2001 and 2008 there were at least 17 deaths involving chatrooms or sites that provide advice on suicide methods.[1716] And in a qualitative study in the UK, one participant reported extensive interaction in pro-suicide chatrooms while looking for encouragement and advice on methods of suicide (including, on one occasion, joining a virtual suicide pact).[1717]

Commenting on content

15.58 Potentially illegal pro-suicide messaging can be found in the comments sections on posted content. An Australian study which looked at comment replies to suicide-related posts on X (formerly Twitter) found that the nature of these replies was often mixed. While some of the comments were helpful, discouraging the suicide attempt or providing support, the study found that almost one in four replies were "dismissive and pro-suicide."[1718]

15.59 Similarly, a study looked at the comment threads of 26 livestreaming videos where an individual had threatened to take their own life.[1719] In nearly 9 out of 10 cases (88%), the study found that comments attempted to discourage the suicide threat, but it also found that in just under half (11 of the 26 cases), some of the comments encouraged the suicide

---

[1714] Phillips, J G., Diesfeld, K. and Mann, L., 2019. Instances of online suicide, the law and potential solutions, *Psychiatry, Psychology and Law,* 26 (3). [accessed 27 January 2023].
[1715] Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., Purdy, N., Daine, K. and John, A., 2017. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown, p.16, *PLoS ONE,* 12 (8). [accessed 10 July 2023].
[1716] Cohen-Almagor, R, and Lehman-Wilzig, S., 2022. Digital Promotion of Suicide: A Platform-Level Ethical Analysis, p.6, *Journal of Media Ethics,* 32 (2). [accessed 10 July 2023].
[1717] Biddle, L. et al, 2018.
[1718] O'Dea, B., Achilles, M R., Larsen, M E., Batterham, P J., Calear, A L. and Christensen, H., 2021. The rate of reply and nature of responses to suicide-related posts on Twitter. p.2, *Internet Interventions*, 13. [accessed 11 July 2023].
[1719] Videos analysed took place between 2001 and 2017 and included cases from the USA, India, UK, France, Turkey, Canada, Russia, Sweden, Japan and Thailand.

attempt or insulted the victim. Audience anonymity was cited as one of the potential factors contributing to this online baiting behaviour.[1720]

15.60 A report which reviewed content on TikTok described various types of suicide-related videos available on the service. One video, providing tips on suicide methods, had 24,000 views and over 200 comments, with many comments providing specific advice and information on suicide methods, including the use of common household items.[1721]

Posting content (text, images, videos)

15.61 The ability to post content is an important functionality mentioned in the research and literature on suicide and self-harm. It enables users to communicate and establish contact with others who are experiencing similar thoughts or behaviours, but the evidence shows they it is also being used to negatively influence users' thinking around suicide.

15.62 A UK-based qualitative study (where participants had either previously used the internet for suicide-related purposes or had been admitted to hospital following serious self-harm) found that among the self-harm patients, most had avoided generating online dialogue and instead preferred to observe others' posts. The study explained that almost all had viewed others' posts on online services about self-harm methods and had used these as a source of information that they could search to gain insight into experiences with these methods, or to decide on details of implementation.[1722]

15.63 A small-scale qualitative study in the UK looking at 18 to 24-year-olds who had previously had suicidal thoughts, found that most participants in the study reported that they had used the internet to communicate with others about their suicidal feelings, which most had found offered them a strong and supportive sense of community.[1723] However, some noted pessimistic posts as the main factor generating negative effects.

15.64 Another paper identified that graphic images and videos posted online were, in some cases, found to be emotionally disturbing by people with a history of self-harm and "*potentially triggering of self-harm behaviour.*"[1724] The studies covered various types of posts such as images of wounds and scars, self-harm memes, videos with NSSI (non-suicidal self-injury) content, suicide images from a first-person and third-person perspective, and content containing images of self-harm coupled with negative words (such as 'suicide' and 'death').[1725]

[1720] Phillips, J G. and Mann, L., 2019. Suicide baiting in the internet era p.1, *Computers in Human Behaviour,* 92. [accessed 11th July 2023].
[1721] Ekō, 2023. Suicide, Incels, and Drugs: How TikTok's deadly algorithm harms kids p.10. [accessed 11 July 2023].
[1722] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.12, *PLoS ONE,* 13 (5). [accessed 10 July 2023].
[1723] Bell, J., Mok, K., Gardiner, E. and Pirkis, J., 2017. Suicide-related internet use among suicidal young people in the UK: Characteristics of users, effects of use, and barriers to offline help-seeking, pp.11-12, *Archives of suicide research: official journal of the International Academy for Suicide Research*, 22 (4). [accessed 10 July 2023].
[1724] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023. Research Review: Viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms p.17, *Journal of Child Psychology and Psychiatry,* 64 (8). [accessed 10 July 2023].
[1725] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023. Research Review: Viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms, pp.4-11, *Journal of Child Psychology and Psychiatry,* 64 (8). [accessed 10 July 2023].

15.65 The ability to save and return to content that has been posted may also increase the risk of harm due to the potential to increase the frequency of exposure, and increase the risk presented by highly depressive content that becomes harmful when viewed in large quantities over time.[1726]

Reacting to content and re-posting or forwarding content

15.66 Validation from other users on a service, through means such as 'likes', comments or re-posting, can reinforce or even exacerbate negative thought patterns or behaviours, and potentially encourage the further posting of potentially harmful content. It can also provide users with a sense of community in feeling that they are not alone in their thinking.[1727]

Livestreaming

15.67 A livestreaming functionality can increase the risk of users being exposed to suicide content. It can also increase the risk of those hosting the livestream, who may be struggling themselves, being exposed to harmful messages. Notably, livestreaming intersects with group messaging and commenting functionalities – users can often message one another as a group within the livestream or leave comments. While some users may use these messages or comments to express sympathy or coordinate help, some can encourage suicide or serious self-harm.

15.68 There have been numerous cases of livestreaming functionalities being used to show users self-harming or ending their life in real time. In 2008, a 19-year-old male publicly took his own life while livestreaming. Viewers were able to watch this happen in real time, with some viewers encouraging him to continue the attempt.[1728]

15.69 A research paper looking specifically at evidence related to a social media service and suicidal behaviour identified that those who livestreamed suicidal behaviour were mainly under 35 years old, and the majority were male.[1729]

Content exploring

Content tagging

15.70 The ability to tag content so that other users can find it and content similar to it, such as by using hashtags, is also a risk factor for disseminating suicide-related content. Variations of suicide and self-harm-related hashtags may be used to avoid content removal. These hashtags can often create spaces where harmful content can proliferate for extended periods without detection by online services. The use of some hashtags to disguise the true

---

[1726] Molly Rose Foundation and The Bright Initiative. 2023. Preventable yet pervasive: The prevalence and characteristics of harmful content, including suicide and self-harm materials, on Instagram, TikTok and Pinterest. [accessed 10 October 2024]; BBC News, 2022. Molly Russell: Social media causes no end of issues, head says. [accessed 10 October 2024].

[1727] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.12, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[1728] Cohen-Almagor, R, and Lehman-Wilzig, S., 2022. Digital Promotion of Suicide: A Platform-Level Ethical Analysis, p.2, *Journal of Media Ethics,* 32 (2). [accessed 10 July 2023].

[1729] Shoib, S., Chandradasa, M., Nahidi, M., Amanda, T W., Khan, S., Saeed, F., Swed, S., Mazza, M., Di Nicola, M., Martinotti, G., Di Giannantonio, M., Armiya'u, A Y. and De Berardis, D., 2022. Facebook and Suicidal Behavior: User Experiences of Suicide Notes, Livestreaming, Grieving and Preventive Strategies - A Scoping Review, p.8, *International Journal of Environmental Research and Public Health,* 19. [accessed 11 July 2023].

nature of suicide and self-harm content may also increase the risk that more users will unintentionally encounter this content.[1730]

15.71 In other cases, variations of hashtags that are likely to be blocked have been used to continue to access the content. Despite some effort by services to remove hashtags associated with suicide and self-harm, research conducted by the Molly Rose Foundation in 2023 found that it was still possible to retrieve large amounts of potentially harmful suicide and self-harm related content on Instagram.[1731]

User-generated content search filtering

15.72 In some cases, users can apply filters when they search for user-generated content (UGC) on a U2U service to remove or avoid supportive content (for example, links to a support service) that may support the user who is having suicidal or self-harm thoughts.

15.73 Participants in a UK research study, who had either previously used the internet for suicide-related purposes or had been admitted to hospital following serious self-harm, described how they would 'sift through' UGC. The study found that some participants would actively avoid or block out online help, such as pop-ups and support links, once their suicidal thoughts became intense and would filter user-generated data to remove support-giving responses.[1732]

Hyperlinking

15.74 Hyperlinks may contribute to the risks of harm related to suicide and self-harm content. Some studies have shown that hyperlinks can cause a 'rabbit-hole' effect, whereby users engage with links to similar content, leading them to more harmful content which they had not necessarily set out to view.[1733]

15.75 A study on suicide-related internet use found that many young adults in the sample followed links within and across different online services. The study found that this behaviour tended to increase as mood lowered, leading to an escalation in browsing and exposure to issues that the participants had not previously considered.[1734]

Recommender systems

Content recommender systems

15.76 Some evidence suggests that content recommender systems[1735] can increase the risk of exposure to suicide-related content. As recommender systems are understood to maximise

---

[1730] Arendt, F., Scherr, S. and Romer, D., 2019. Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults, p.3, *New Media & Society,* 21 (11-12). [accessed 10 July 2023].

[1731] Molly Rose Foundation and The Bright Initiative. 2023. Preventable yet pervasive: The prevalence and characteristics of harmful content, including suicide and self-harm materials, on Instagram, TikTok and Pinterest. [accessed 10 October 2024].

[1732] These participants said that by this point, they had decided that they wanted to end their life and were online to research how to action it, looking only for this type of user-generated content. Source: Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.11, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[1733] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.8, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[1734] Biddle, L. et al, 2018.

[1735] In the context of online services, a recommender system (or a recommender engine) is a type of information retrieval and ranking system that curates content to a service user. Recommender systems are powered by a set of algorithms

user engagement, they can make it more likely that users who engage with harmful content see more of it in the future. In a national survey by Swansea University and Samaritans (where 87% of the sample reported having self-harmed before), more than four in five (83%) respondents reported coming across self-harm and suicide content through feeds of recommended content on social media, despite not having searched for it.[1736] Recommender systems can recommend potentially harmful suicide related content very soon after a new user first signs on. Researchers from the Centre for Countering Digital Hate in the USA created four 'standard' new accounts with a female username on TikTok for users aged 13 in the USA, the UK, Australia and Canada. Four separate accounts were created with a username that indicated a body image-related concern.[1737] The researchers found that the 'standard' teen TikTok accounts recommended self-harm, suicide and eating disorder content within minutes of scrolling the 'for you' feed. Suicide content appeared within the first 2.6 minutes.[1738] The limited sample of this study means that we cannot be confident that the findings are representative of a realistic user experience on the platform. Nonetheless, this research does demonstrate that it is *possible* for this content to be served up within a very short period of the first viewing session.

15.77 It has been noted that recommender systems – or "the use of algorithms to provide content" – can play a role in enabling or encouraging episodes of binging on large volumes of potentially harmful content, a factor that has been highlighted by the Coroner's Service to have contributed to deaths in the UK.[1739] While pieces of content judged in isolation may not be considered illegal, such instances demonstrate the potential cumulative effect and risks of harm amounting from sustained exposure to suicide and self-harm-related content propagated by recommender algorithms.[1740]

15.78 Dr Ysabel Gerrard, a member of the Facebook and Instagram Suicide and Self-Injury (SSI) Advisory Board, stated: "In particular, it's important that we pay attention to platforms' recommender algorithms (the process of showing users more content they might want to see). People who are already viewing content about self-harm, eating disorders and suicide are likely to get it recommended back to them, and we don't know enough about the role this algorithmic process might play in their mental ill health."[1741] Recommended search terms or search term completions can also exacerbate risk by amplifying prior tendencies to view harmful suicide related content by reducing friction in the search process or

---

which, depending on what they are optimised for, set the decision path for what content is suggested to the user. The goal of a recommender system is to generate recommendations likely to engage the user, although the exact metric/goal will vary by platform.

[1736] Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. p.4. [accessed 10 July 2023].

[1737] Across all accounts, researchers expressed an interest in body image, mental health and eating disorders by watching and liking relevant videos.

[1738] Centre for Countering Digital Hate, 2022. Deadly By Design: TikTok pushes harmful content promoting eating disorders and self-harm into users' feeds p.19. [accessed 11 July 2023].

[1739] The Coroner's Service, 2022. Prevention of Future Deaths. [accessed 10 October 2024].

[1740] The Coroner's Service, 2022.

[1741] University of Sheffield (Dr Ysabel Gerrard), How we're helping social media companies remove harmful content and protect their users. [accessed 10 January 2023].

introducing users to novel search terms that may surface more harmful content that the user would not otherwise have seen.[1742]

Risk factors: Business models and commercial profiles

15.79 There is some evidence to suggest that advertising-based revenue models may be a risk factor for suicide and self-harm content. In its 2023 Protection of Children Call For Evidence (CFE) response, the Molly Rose Foundation noted that email and push notifications can direct children to suicide and self-harm content. These are sent to users to encourage continued engagement with a service provider to drive up advertising revenue, increasing the risk by encouraging a user to revisit potentially harmful recommended content that the user may have previously engaged with.[1743] Some evidence suggests that there are instances where this revenue model can suggest further suicide and self-harm content to an online user.[1744]

[1742] For example, after simulating a proclivity for viewing suicide and self-harm related content, The Molly Rose Foundation discovered that avatar accounts on TikTok were served up problematic recommended search terms "people also search for 'quickest way to end it', 'I am going to end it soon', 'I am going to end it'", and autocompletions for the prompt string "want to" which included "end it", "cut", "give up" and "go missing". Source: Molly Rose Foundation and The Bright Initiative. 2023. Preventable yet pervasive: The prevalence and characteristics of harmful content, including suicide and self-harm materials, on Instagram, TikTok and Pinterest. [accessed 10 October 2024].

[1743] Molly Rose Foundation response to 2023 Ofcom Call for Evidence.

[1744] One example provided was an email sent to Molly Russell before she took her own life. The Call For Evidence response states that this email contained images of self-harm (some of a graphic nature), suicide (including methods) and depression. Source: Molly Rose Foundation response to 2023 Ofcom Call for Evidence.

16. Foreign interference offence

Summary analysis for the foreign interference offence: how harm manifests online and risk factors

The new Foreign Interference Offence (FIO) has been designed to tackle malign activity carried out for, or on behalf of, or intended to benefit, a foreign power. Prohibited conduct captured by this offence will include where there is a misrepresentation of a person's identity or purpose, or in the presentation of the information, for example, through state-backed disinformation campaigns.

In introducing this new offence, the UK Government has explained that: "Foreign interference is intended to sow discord, manipulate public discourse, discredit the political system, bias the development of policy, and undermine the safety or interests of the UK".

Harm that can arise from this offence is wider than the individual and can affect societies as a whole. For example, a foreign state could seek to manipulate whether or how someone participates in an electoral event through state-sponsored disinformation campaigns. This would have implications on the country's electoral outcomes, undermining the integrity of elections and creating mistrust in online information.

The rapid pace of development of generative AI models and technology has been recognised as posing a risk which could be exploited by those engaging in foreign interference.

Generative AI models and technology present the opportunity for perpetrators to create an increased volume of foreign interference content, with increased quality and personalisation of content for targeted audiences, as well as reducing costs and barriers to entry for perpetrators of foreign interference campaigns.

At present, it appears that generative AI technologies are more likely to significantly exacerbate existing risks of foreign interference and other information-based threats, rather than present wholly new risks.

Service type risk factors:

There is a particular risk of FIOs happening on **social media services**, where perpetrators of the offence can create fake profiles which can be manipulated by bots.

There is evidence of FIO and influence operations[1745] occurring across many different service types, using different tactics. These services include information-

[1745] The Carnegie Endowment for International Peace defines influence operations as *"**organized attempts to achieve a specific effect among a target audience**. Such operations encompass a variety of actors—ranging from advertisers to*

sharing services, discussion forums and chat rooms, and private messaging services. While we know more about how these operations are carried out on certain services, this is not necessarily an accurate reflection of the presence of the harm. More attention and resources have likely been devoted to studying influence operations on some services than on others due to availability of data for study.

User base risk factors:

Foreign influence operations have previously targeted **personal characteristics** such as race, religion, sexuality, and gender. Foreign interference operations often exploit and build on narratives that are common in society, including narratives that negatively depict, or target people based on their personal characteristics. For example, previous influence operations have targeted female politicians to spread and amplify gendered narratives and expectations that undermine them, increasing the risks of harm to women from foreign influence operations that amplify these pre-existing narratives. Evidence also suggests that diaspora groups and those who hold intersectional identities may be disproportionately at risk of harm from foreign influence operations.

Functionalities and recommender systems risk factors:

Some functionalities might increase the likelihood that influence operations will be encountered by users, thereby increasing the risk of harm.

The ability to create **fake user profiles** can be exploited by perpetrators of foreign interference operations – both to disseminate content and to impersonate authoritative and high-profile sources. The use of coordinated networks on social media accounts can also be used to amplify content and spread narratives across services. The functionality of **user connections** is therefore a risk factor for this offence.

Services where users can more easily share this content, both within and across services, are particularly risky. This is because they enable foreign influence operations to spread between services and other online spaces, thereby broadening their effect. These functionalities include **re-posting** and **forwarding content, encrypted messaging,** and mechanisms for sharing information across services, such as the use of **hyperlinks**.

Posting from **anonymous user profiles** can also be used in foreign interference operations and to spread disinformation on services, as well as the ability to **post content**, especially types of content that combine images or videos and text.

Service **recommender algorithms** can also increase the risks of harm from foreign influence, as they tend to amplify content with high user engagement. Potential perpetrators can therefore manipulate these algorithms to spread harmful content

activists to opportunists—that employ a diverse set of tactics, techniques, and procedures to affect a target's decision making, beliefs, and opinions". Source: Carnegie Endowment for International Peace (Thomas, E., Thompson, N., and Wanless, A.), 2020. The Challenges of Countering Influence Operations. [accessed 11 September 2023].

more widely by reposting selected content or coordinating the mass sharing of harmful content. This also allows bad actors to increase user exposure to foreign interference content for the intended purpose of manipulating or misleading users.

Business model risk factors:

Services which raise income through **advertising** may be exploited by potential perpetrators who can use advertisements as an opportunity to spread foreign interference content. This will be more effective if it allows them to target specific segments of the population with their adverts, without identifying the funder of the advertising.

Introduction

16.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to the Foreign Interference Offence (FIO) listed under 'Relevant offences'; and
- The use of these services for the commission and/or facilitation of this offence (collectively, the 'risks of harm').

16.2 We set out the characteristics of U2U services that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual or encountered in combination with content of a different kind.

16.3 The FIO is a new criminal offence set out in section 13 of the National Security Act 2023.

16.4 As this is a new offence, there are no legal precedents.[1746] In preparing this risk assessment, we have therefore considered evidence of conduct that appears to be broadly aligned with the conduct that is intended to fall within the scope of the new offence. Using the evidence as a proxy, we have drawn inferences about the characteristics of services that may be relevant to the risk of this new offence. We will keep our evidence base under review as new evidence emerges.

16.5 Attribution is a key challenge associated with the identification of activities that may constitute an FIO. This is increasingly difficult due to the use of commercial bot networks, digital marketing companies and local content creators which obscure the involvement of state actors in influence operations and make it harder to conclusively attribute operations.[1747] For example, the recent US Department of Justice indictments revealing the

[1746] The FIO is a conduct-based offence. The forms of conduct that this offence can fall under are varied and diverse – they count as part of committing the offence if they meet the three conditions required for the offence to be present. This conduct can include diverse tactics, including creating an account on a social media platform impersonating a British politician to post content in support of the interests of a particular nation state, and posting memes with deliberate and strategic intent to sway public opinion in the UK on behalf of another, hostile nation.

[1747] Carnegie Endowment for International Peace (Thomas, E., Thompson, N., and Wanless, A.), 2020. The Challenges of Countering Influence Operations. [accessed 11 September 2023].

alleged funding of a Tennessee-based online content creation company by Russian operatives to create and distribute content to American audiences with hidden Russian government messaging as part of an influence operation shows how local influencers and content creators can be used to obfuscate the role of state actors in influence operations.[1748] In addition, conflicting motives have made attribution more difficult; for example, when influence campaigns covertly carried out by state-linked operatives generate significant financial gains for the perpetrators, platforms may focus on the commercial aspects and miss the state coordination behind the activity.[1749] Recent research has also revealed evidence of inter-state coordination within state-backed information operations, which may further complicate attribution attempts.[1750]

16.6 Targeted governments may not wish to publicly attribute operations to foreign powers. This is especially likely when the explicit purpose of an influence operation is not to be publicly attributed to the sponsoring state power. A notable example of this is the Internet Research Agency's[1751] attacks during the US 2018 midterm elections.[1752]

16.7 The evidence focuses on confirmed foreign interference and influence operations on U2U services. Throughout this chapter, 'foreign interference', 'foreign influence' and 'information operations' will be used interchangeably to refer to conduct that Ofcom considers likely to meet the conditions outlined in the offence.

Relevant offences

16.8 The Act requires Ofcom to consider the risks of harm connected with specific offences. Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Online Safety Act (the Act), which includes the FIO.

16.9 The FIO makes it illegal for a person to engage in conduct for, on behalf of, or with intent to benefit a foreign power, in a way which has or is intended to have an interference effect – for example, to interfere with how a person participates in political or legal processes, to

[1748] US Department of Justice, 2024. Two RT Employees Indited for Covertly Funding and Directing U.S. Company that Published Thousands of Videos in Furtherance of Russian Interests. [accessed 18 October 2024].
[1749] Carnegie Endowment for International Peace (Thomas, E., Thompson, N., and Wanless, A.), 2020. The Challenges of Countering Influence Operations. [accessed 11 September 2023].
[1750] Wang, X., Li, J., Srivatsavaya, E. and Rajtmajer, S., 2023. Evidence of inter-state coordination amongst state-backed information operations, *Scientific Reports*, 13, 7716. [accessed 26 September 2023].
[1751] The Internet Research Agency is a Russian organisation based in St Petersburg that was funded by Yevgeniy Viktorovich Prigozhin and companies he controlled, which conducted social media operations targeted at large US audiences with the goal of sowing discord in the US political system. Source: Mueller, R. S., 2019. Report on the investigation into Russian interference in the 2016 Presidential Election, Volumes I & II. *US Department of Justice Publications and Materials*. 47. [accessed 11 September 2023]. Following the death of Prigozhin in 2023, the future and ownership of the Internet Research Agency is unclear. In June 2023, before his death but after his attempted mutiny against the Russian state, Russian media reported that the agency had been disbanded. Source: Reuters, 2023. Prigozhin-controlled Russian media group shuts after mutiny, 2 July. [accessed 16 May 2024]. However, in March 2024, Google Cloud's Mandiant reported that the infrastructure for covert information operation threat activity from Prigozhin-associated entities remained viable for use: Mandiant, 2024. Life After Death? IO Campaigns Linked to Notorious Russian Businessman Prigozhin Persist After His Political Downfall and Death. [accessed 16 May 2024].
[1752] Francois, C. and Douek, E., 2021. The Accidental Origins, Underappreciated Limits, and Enduring Promises of Platform Transparency Reporting About Information Operations, *Journal of Online Trust and Safety*, pp.1-30. [accessed 27 September 2023].

interfere with the exercise of public functions or prejudice the safety or interests of the UK.[1753]

16.10 For the FIO to be committed, three conditions must be met. In summary:

a) the conduct is 'prohibited conduct' (for example, it constitutes an offence or involves misrepresentation or coercion);
b) there is a link between the conduct and a foreign power (such as a foreign government); and
c) the conduct, or course of conduct, is intended to have a certain effect or objective (the interference effect).

16.11 Misrepresentation and coercion of any kind are two primary types of conduct covered by the FIO.[1754]

16.12 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of the offence (and in relation to offences in Scotland, being involved art and part in the commission of these offences).

16.13 For more information on the offence and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance or ICJG.

16.14 Foreign interference is highly context-dependent and varies across locations, services and target audiences. It relies on determining whether the behaviour or action constitutes an offence, rather than the type of content itself. The offence can be carried out by an individual or a group of people. It can also be carried out where a person is reckless as to whether the prohibited conduct or a course of conduct of which it forms part, will have an interference effect.

16.15 Foreign influence is an area of focus for the UK's security services.[1755] When introducing the offence, the UK Government outlined that its principal aim was to create a more challenging operating environment for, and to deter and disrupt the activities of, foreign states that seek to undermine the UK's interests, institutions, political systems and rights, or prejudice the UK's national security. The FIO will also seek to protect diaspora communities from the influence of foreign powers.[1756]

16.16 The UK Government has explained that the FIO will be a tool for deterrence and disruption and raise the cost of carrying out interference activities that target the UK,[1757] particularly in elections.

16.17 There is no 'generic' example of what foreign interference looks like.[1758] We know that operations are often intended to sow discord and frequently target sensitive events like

---

[1753] Sections 13 and 14 of the National Security Act 2023.
[1754] Section 15 of the National Security Act 2023.
[1755] During his annual threat update in November 2022, MI5 Director General Ken McCallum highlighted that Russia's covert actions targeting the UK include disinformation and democratic interference, and he highlighted ongoing threats to Chinese diaspora members and Iranian dissidents made by their respective regimes in the UK. Source: McCallum, K., 2022. Annual Threat Update. [accessed 11 September 2023].
[1756] An example provided on where this might be needed is when an individual is threatened because of their views on a foreign power's foreign policy. Source: House of Lords. National Security Bill (parliament.uk)
[1757] Home Office, 2023. Foreign interference: National Security Act factsheet. [accessed 25 January 2023].
[1758] In the Explanatory Notes for the National Security Act, the UK Government has provided the following example of how the offence may include online conduct. A foreign power runs a covert unit of state actors operating a troll farm, an

elections, referendums and health emergencies, among others. For example, the interference operation run by Russia's Internet Research Agency during the 2016 US Presidential election, is by far the most studied and from which we draw on in this chapter.

16.18 There are also some examples of events generally considered to have been subject to foreign influence in the UK. These include the compromise of US to UK trade documents that were leaked ahead of the 2019 UK General Election,[1759] [1760] [1761] Russian influence operations surrounding the attempted assassination of Sergei Skripal,[1762] a network of People's Republic of China-linked accounts amplifying the activities of Chinese diplomats in the UK online,[1763] and Russian influence operations following its invasion of Ukraine in February 2022.[1764]

How foreign interference manifests online

16.19 This section is an overview which looks at how FIO offences manifest online, and how individuals may be at risk of harm.

16.20 By their nature, foreign interference activities are typically covert and due to the involvement of nation-states, are often sophisticated.

16.21 As such, establishing the presence of FIO in the UK is challenging. A systematic analysis of reported influence efforts includes four operations targeting the UK, and 76 foreign influence operations across 49 targeted countries.[1765] In his 2024 Annual Threat Update, MI5 Director General Ken McCallum noted that the number of state threats investigations the agency is running has increased by 48% over the past year. He added that, alongside law enforcement partners, MI5 has responded to twenty Iranian-backed plots targeting British citizens and UK residents.[1766]

16.22 There is no robust data about the number of UK internet users exposed to these campaigns, nor the nature of the campaigns' audiences. The evidence we have assessed shows that some foreign interference operations in the UK have targeted people based on their protected characteristics, using conduct that appears to be broadly aligned with the

---

organisation employing people to make deliberately offensive or provocative posts online to manipulate public opinion or cause conflicts via a variety of different tools. The troll farm uses coordinated inauthentic behaviour and online manipulation to create and amplify disinformation on the efficacy and alleged side effects of vaccines for children and uses misrepresentations and false identities to infiltrate legitimate debates on the topic. Through these actions, the foreign power aims to undermine the use of public health services by amplifying an existing 'wedge' issue to disrupt social cohesion. Source: Home Office, 2022. National Security Bill: Explanatory Notes. [accessed 27 September 2023].

[1759] Wendling, M., 2019. General election 2019: Reddit says UK-US trade talks document leak 'linked to Russia, *BBC News*, 7 December. [accessed 27 September 2023].

[1760] Graphika (Nimmo, B.), 2019. 'UK Trade Leaks'. [accessed 12 September 2023].

[1761] National Cyber Security Centre, 2023. UK and allies expose Russian intelligence services for cyber campaign of attempted political interference. [accessed 9 July 2024].

[1762] Global Engagement Centre, 2022. GEC Special Report: The Kremlin's Chemical Weapons Disinformation Campaigns. [accessed 12 September 2023].

[1763] Schliebs, M., Bailey, H., Bright, J. and Howard, P. N., 2021. People's Republic of China's Inauthentic UK Twitter Diplomacy: A Coordinated Network Amplifying PRC Diplomats, Programme on Democracy & Technology. [accessed 11 September 2023].

[1764] Foreign, Development and Commonwealth Office, 2022. UK Exposes Sick Russian Troll Factory plaguing Social Media with Kremlin Propaganda. [accessed 27 September 2023].

[1765] Martin, D. A., Shapiro, J. N. and Ilhardt, J., 2020. Trends in Online Influence Efforts, Empirical Studies of Conflict Project, 2. [accessed 27 September 2023].

[1766] McCallum, K., 2024. Director General Ken MacCallum gives latest threat update. [accessed 17 October 2024].

conduct intended to fall within the scope of the FIO, more details of which are discussed in the Risk Factors: User Base section.

16.23 In the Explanatory Notes for the National Security Act, the UK Government has provided the following example of how the offence may include online conduct; A foreign power runs a covert unit of state actors operating a troll farm, an organisation employing people to make deliberately offensive or provocative posts online to manipulate public opinion or cause conflicts via a variety of different tools. The troll farm uses coordinated inauthentic behaviour and online manipulation to create and amplify disinformation on the efficacy and alleged side effects of vaccines for children, and uses misrepresentations and false identities to infiltrate legitimate debates on the topic. Through these actions, the foreign power aims to undermine the use of public health services by amplifying an existing 'wedge' issue to disrupt social cohesion.[1767]

16.24 Foreign influence operations online are conducted using a variety of tactics. Disinformation is one of the most frequently reported and is often associated with FIO. It is an overarching tactic often used in foreign influence operations. Disinformation is strongly linked to the misrepresentation elements contained in the FIO. More specific tactics are used on U2U services, often as part of broader foreign influence operations.

16.25 Tactics on U2U services can include coordinated behaviour, cross-platform coordination and inauthentic behaviour, manipulation or impersonation of journalism, amplification of conspiracy narratives, the creation of disinformation, deepfakes or cheap-fakes, and the use of automated bots. Other activities that could constitute a foreign interference tactic include publishing individuals' private or identifiable information, hack and leak operations, and distributed denial of service (DDoS) attacks.

16.26 The rapid pace of development of generative AI models and technology has been recognised as both a risk and an opportunity, with the World Economic Forum's Global Risks Report 2024 highlighting how AI can amplify manipulated and distorted information that can destabilise societies.[1768] At present, it appears that generative AI technologies have the potential to significantly exacerbate existing risks of foreign interference and other information-based threats, rather than present wholly new risks. For example, large language models and other types of AI can reduce the amount of time and costs involved in conducting foreign interference campaigns.[1769]

16.27 Based on analysis of the risks generative AI poses to the proliferation of misinformation, we can consider the risks of generative AI as falling into four interlinked categories:

- Increased quantity of foreign interference content
- Increased quality of foreign interference content
- Increased personalisation of foreign interference content[1770]

[1767] Home Office, 2023. National Security Act: Explanatory Notes. [accessed 13 August 2024]
[1768] World Economic Forum, 2024. Global Risks Report 2024. [accessed 16 May 2024].
[1769] Brookings Institute (Brandt, J.), 2023. Propaganda, foreign interference and generative AI. [accessed 16 May 2024]
[1770] Simon, F. M., Altay, S., and Mercier, H., 2023. Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown, *Harvard Kennedy School Misinformation Review*, 3 (1). [accessed 16 May 2024].

- Lower costs and reduced barriers to entry for perpetrators of foreign interference campaigns

16.28 These four interlinked categories of risk suggest that generative AI tools and technologies could be used by perpetrators of the foreign interference offence to target hyper-specific or niche audiences with personalised content.

16.29 Across the world, there has been evidence of the use of AI-generated audio, image, video and text-based content, by both state-linked and non-state actors, to influence elections and public opinion more broadly. The ability to create misleading and polarising content with an increased level of personalisation and significantly reduced costs, is frequently highlighted as a particularly pertinent concern.[1771] A study has found that AI models are better at both producing accurate information that is easier to understand, and at producing more compelling disinformation, than humans. The same study found that participants were not able to distinguish between social media posts generated by an AI model and social media posts produced by humans.[1772] Due to this, the increasing availability and sophistication of generative AI technologies and tools mean that we could see an increase in the amount of foreign interference attempts, and a potential increase in the success of these attempts.

16.30 Influence operations are increasingly cross-border and cross-service, with campaigns that involve similar content spread across country-specific distribution lists and networks, pushing the same agenda across targeted nations.[1773] Such campaigns have targeted the UK,[1774] and evidence shows that Russia, the People's Republic of China, Saudi Arabia and the UAE have all engaged in cross-jurisdictional operations.[1775] For example, Google's Threat Analysis Group reported that it had disrupted over 50,000 instances of activity from Chinese-linked information operation DRAGONBRIDGE (also referred to as 'Spamouflage Dragon') across YouTube, Blogger and AdSense in 2022. This was the most prolific information operation the group had tracked.[1776]

16.31 The evidence we have assessed suggests that the deployment of bots[1777] can be exploited by perpetrators of the FIO. Under the direction of a person, they can generate or amplify content as part of foreign influence operations. Researchers have also suggested that generative AI can enable foreign interference campaigns to deploy social media bots with the ability to participate in conversations, with the potential to engage voters with

---

[1771] Brookings Institute (Brandt, J.), 2023. Propaganda, foreign interference and generative AI. [accessed 16 May 2024].
[1772] Spitale, G., Biller-Andorno, N. and Germani, F., 2023. AI model GPT-3 (dis)informs us better than humans, *Science Advances*, 9 (26). [accessed 21 October 2024].
[1773] Martin, D. A., Shapiro, J. N. and Ilhardt, J., 2020. Trends in Online Influence Efforts, Empirical Studies of Conflict Project, 2. [accessed 27 September 2023].
[1774] Carnigie Endowment For International Peace (Thomas, E., Thompson, N. and Wanless, A.), 2020. The Challenges of Countering Influence Operations. [accessed 27 September 2023].
[1775] Martin, D. A., Shapiro, J. N. and Ilhardt, J., 2020. Trends in Online Influence Efforts, Empirical Studies of Conflict Project, 2. [accessed 27 September 2023].
[1776] Threat Analysis Group (Butler, Z. and Taege, J.), 2023. Over 50,000 instances of DRAGONBRIDGE activity disrupted in 2022. [accessed 27 September 2023].
[1777] 'Bots' is an umbrella term that refers to a software application or automated tool that has been programmed by a person to carry out a specific or predefined task without any human intervention.

convincing, and often increasingly personalised, information designed to polarise or mislead.[1778] [1779]

16.32 Bots are typically employed on social media services to simulate human behaviour. They are often used in foreign influence operations and can be used for like[1780] and click farming,[1781] hashtag hijacking,[1782] initiating a repost storm (when a post is instantly reposted by a network of accounts) and trend-jacking.[1783] [1784] Research from the Oxford Computational Propaganda Research Project found that in 2020, 57 countries used bots or automated accounts as part of their efforts to influence the online sphere.[1785],[1786]

16.33 While most research on foreign influence operations focuses on three threat actors (Russia, the People's Republic of China and Iran), foreign influence campaigns originate from across the globe.

Risks of harm to individuals presented by the foreign interference offence

16.34 Overall, the evidence available does not allow us to draw comprehensive conclusions about the impact and harm to individuals associated with foreign influence operations. However, as outlined in this chapter, we are aware that such activity can potentially simultaneously target and impact UK individuals and society. Such operations are also a significant source of concern for the population, policymakers and security services. Research that identifies how operations have influenced people and societies by "*altering beliefs, changing voting behaviour, or inspiring political violence – is limited and scattered*".[1787] Despite the lack of direct insight, there is clear potential that there is a risk of harm to individuals which underpins the desire by the UK Government to make the FIO a priority offence in the online safety regime.

16.35 Disinformation, including where used in foreign interference operations, can have broader aims than influencing the outcome of an individual election. The Intelligence and Security

---

[1778] Brookings Institute (Brandt, J.), 2023. Propaganda, foreign interference and generative AI. [accessed 16 May 2024]
[1779] Goldstein, J. A., Sastry, G., Musser, M., DiResta, R., Gentzel, M., and Sedova, K., 2023. Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations, *Georgetown University's Center for Security and Emerging Technology, OpenAI and Stanford Internet Observatory*. [accessed 21 October 2024].
[1780] 'Like farming' refers to the use of fake pages on social media services designed to artificially increase the popularity of a page, so it can be sold to buyers seeking accounts with large followings or for scam and fraud activity.
[1781] 'Click farming' refers to the practice of manually clicking on online adverts to increase the clickthrough rate value, boost engagement metrics, and inflate impressions. This activity can be carried out by bot accounts, as discussed here, or by large groups of workers.
[1782] 'Hashtag hijacking' refers to the use of a hashtag for a purpose other than it was created – such as tagging a message containing undesirable or harmful content with a popular, but unrelated, hashtag to surface this content to a target audience.
[1783] 'Trend jacking' refers to when influencers, brands or organisations insert themselves into conversations online that are gaining a lot of attention – for example, by using associated hashtags or trending audios.
[1784] US Department of Homeland Security, 2018. Social Media Bots Overview. [accessed 13 September 2023].
[1785] Programme on Democracy & Technology (Bradshaw, S., Bailey, H. and Howard, P. N.), 2021. Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation. [accessed 13 September 2023].
[1786] It is unclear whether some of these bots employ GenAI technologies, but we think that GenAI bots could be used in a similar manner.
[1787] Carnegie Endowment For International Peace (Bateman, J., Hickok, E., Courchesne, L.,Thange, I. and Shapiro, J.), 2021. Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research. [accessed 27 September 2023].

Committee's Russia report notes that the aims of these campaigns can include creating an environment of distrust, casting doubt on the true account of events, fomenting political extremism and 'wedge issues', and generally discrediting 'the West'.[1788] In its report on foreign interference in EU democratic processes, the Authority for European Political Parties and European Political Foundations notes that the perception of electoral outcomes being influenced by foreign actors – regardless of effectiveness of any actual interference – erodes public trust in the democratic process of a targeted country, and subsequently undermines the credibility of elected representatives.[1789]

16.36 Bad actors can also attempt to create a perception of delegitimisation and distrust in society and events by creating or amplifying concerns about their own activity. This is an additional risk of foreign interference campaigns, sometimes referred to as 'perception hacking', where malicious actors attempt to play on collective expectations of wide-spread influence and interference operations, hoping to create the perception that they are more impactful and pernicious than they may be.[1790] This process may thus create an overblown sense of concern around foreign interference operations, and exaggerate the impact and effect of otherwise small and ineffective attempted foreign interference campaigns, eroding trust in the target state more than the operation warrants.

16.37 There is some evidence to suggest that gendered disinformation and abuse generated as part of foreign interference campaigns can mean that women are less likely to choose to participate in public life, such as standing for election. Online gendered disinformation and abuse are often intersectional, deploying both sex- and race-based narratives, further exacerbating the threat for women from minority ethnic backgrounds.[1791]

16.38 In addition, some evidence suggests that foreign interference campaigns can target and attempt to influence individuals and public health. In the Explanatory Notes for the National Security Act, the Government explains how a foreign power could utilise a state-sponsored troll farm to create and spread content as part of a foreign interference operation that undermines the use of public health services.[1792] A UK-based study found that one of the most powerful predictors of vaccine hesitancy (a delay in acceptance or refusal of vaccination despite the availability of vaccination services) was conspiracy suspicions[1793], of the sort that might be amplified by the kind of operation described in the Explanatory notes.

16.39 Ofcom's research into the harm caused by deepfakes highlights three primary ways that deepfakes, which are a tactic found in several foreign interference operations, can cause societal and individual harm:

a) Deepfakes that demean: by falsely depicting someone in a specific scenario, for example, engaged in sexual activity.

[1788] Intelligence and Security Committee of Parliament, 2020. Russia. [accessed 16 May 2024].
[1789] Authority for European Political Parties and European Political Foundations, 2023. Foreign Electoral Interference Affecting EU Democratic Processes. [accessed 16 May 2024].
[1790] Meta (Gleicher, N.), 2020. Removing Coordinated Inauthentic Behaviour. [accessed 12 July 2024].
[1791] Wilson Center (Jankowicz, N., Hunchak, J., Pavliuc, A., Davies, C., Pierson, S., and Kaufmann, Z.), 2021. Malign Creativity: How Gender, Sex and Lies are Weaponised Against Women Online. [accessed 16 May 2024].
[1792] Home Office, 2023. National Security Act: Explanatory Notes. [accessed 13 August 2024].
[1793] Allington, D., McAndrew, S., Moxham-Hall, V., and Duffy, B., 2021. Coronavirus conspiracy suspicions, general vaccine attitudes, trust and coronavirus information source as predictors of vaccine hesitancy among UK residents during the Covid-19 pandemic, *Psychological Medicine*, 53, p. 236-247. [accessed 21 October 2024].

b) Deepfakes that defraud: by misrepresenting someone else's identity.
c) Deepfakes that disinform: by spreading falsehoods widely across the internet, to influence opinion on key political or societal issues, such as elections, wars, religion, or health.[1794]

16.40 Media and public concern over foreign influence operations in the UK has also increased since its prominence following the Russian Internet Research Agency's well-publicised influence campaign during the 2016 US Presidential election. In 2020, the British Foreign Policy Group's annual survey found that 27% of Britons saw foreign interference in British politics and democracy as a critical threat, increasing to 32% in 2021.[1795]

Evidence of risk factors on user-to-user services

16.41 We consider that the risk factors below are liable to increase the risks of harm relating to FIO. This is also summarised at the start of the chapter.

Risk factors: Service types

16.42 Research shows that a broad range of services can be used to commit or facilitate offences related to foreign interference. There is significant evidence of influence operations occurring on the following types of services: social media services, video-sharing services, messaging services, information-sharing services, and discussion forums and chat rooms.[1796]

Social media services

16.43 Social media services can be used in various ways in foreign interference campaigns. This includes the use of automated disinformation operations through controlling many fake social media profiles,[1797] as well as bots to simulate human behaviour on social media to disseminate harmful content. As evidenced in the Risk factors: functionalities and recommender systems section below, social media services and video-sharing services have been used in many foreign interference campaigns. These have included a large-scale Russian disinformation operation in the UK and several other countries, targeting Kremlin critics on social media services.[1798]

Information-sharing services

16.44 The evidence shows that some information-sharing services can be exploited by perpetrators of foreign interference operations. A study by the Institute for Strategic Dialogue (ISD) and CASM Technology found several editing-based tactics used by malicious actors on Wikipedia that could make it vulnerable to foreign influence operations; several adversarial edits to Wikipedia's article on the Russo-Ukraine war exhibited narratives

[1794] Ofcom, 2024. A deep dive into deepfakes that demean, defraud and disinform. [accessed 06 September 2024].
[1795] British Foreign Policy Group (Gaston, S. and Aspinall, E.), 2021. UK Public Opinion on Foreign Policy and Global Affairs: Annual Survey 2021. [accessed 27 September 2023].
[1796] While our evidence only names discussion forums, we expect a similar risk of harm to arise from chat room services due to similarities in the characteristics typically found on these service types.
[1797] See 'Risk factors: functionalities and recommender systems' section for more information. Source: Kirchgaessner, S., Ganguly, M., Pegg, D., Cadwalladr, C. and Burke, J., 2023. Revealed: the hacking and disinformation team meddling in elections, *The Guardian*, 15 February. [accessed 15 February 2023].
[1798] See 'Risk factors: functionalities and recommender systems' section for more information. Source: Foreign, Development and Commonwealth Office, 2022. UK Exposes Sick Russian Troll Factory plaguing Social Media with Kremlin Propaganda. [accessed 27 September 2023].

consistent with Russian state-sponsored information warfare.[1799] The Wikimedia Foundation has also banned several editors linked to a group from the People's Republic of China.[1800]

Discussion forums and chat rooms

16.45 Discussion forums have been used in foreign influence campaigns. This includes alleged Russian interference in the 2017 French Presidential election, where disinformation first surfaced on a discussion forum.[1801] 'Online forums' were also used in a North Korean operation in 2010, through fake accounts that disseminated disinformation.[1802.]

16.46 Services that facilitate the building of online communities may also be used by bad actors to target vulnerable individuals with specific characteristics through foreign interference.

Messaging services

16.47 Messaging services are also a risk factor for foreign interference. Research shows that private and encrypted messaging applications are increasingly common channels for foreign influence operations[1803]. The ability to forward content on messaging services has been a primary mechanism for the spread of disinformation, including foreign influence operation content.

16.48 Messaging services popular with certain diaspora communities have become an increasingly common channel for exploitation in foreign influence operations, exploiting users' ability to forward messages.[1804] [1805]

Risk factors: User base

User base demographics

16.49 The following section outlines key evidence of user base demographic factors and risks of harm, which can include personal characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

16.50 Data suggests that user base characteristics including the **gender, religion, sexuality and race** of users could lead to an increased risk of harm to individuals.

[1799] Institute for Strategic Dialogue and CASM Technology (Miller, C., Smith, M., Marsh, O., Balint, K., Inskip, C. and Visser, F.), 2022. Information Warfare and Wikipedia. [accessed 27 September 2023].
[1800] Institute for Strategic Dialogue and CASM Technology (Miller, C., Smith, M., Marsh, O., Balint, K., Inskip, C. and Visser, F.), 2022. Information Warfare and Wikipedia. [accessed 27 September 2023].
[1801] See 'Risk factors: functionalities and recommender systems' section for more information. Source: RAND Corporation (Cohen, R. S., Beauchamp-Mustafaga, N., Cheravitch, J., Demus, A., Harold, S. W., Hornung, J.W., Jun, J., Schwille, M., Tryger, E. and Vest, N.), 2021. Combatting Foreign Disinformation on Social Media: Study Overview and Conclusions. [accessed 27 September 2023].
[1802] RAND Corporation (Cohen, R. S., Beauchamp-Mustafaga, N., Cheravitch, J., Demus, A., Harold, S. W., Hornung, J.W., Jun, J., Schwille, M., Tryger, E. and Vest, N.), 2021. Combatting Foreign Disinformation on Social Media: Study Overview and Conclusions. [accessed 27 September 2023].
[1803] Carnegie Endowment for International Peace (Goodwin, C. and Jackson, D.), 2022. Partnership for Countering Influence Operations, Carnegie Endowment for International Peace. Global Perspectives on Influence Operations Investigations: Shared Challenges, Unequal Resources. [accessed 27 September 2023].
[1804] Nguyễn, S., Kuo, R., Reddi, M., Li, L. and Moran, R. E., 2022. Studying Mis- and Disinformation in Asian Diasporic Communities: The Need for Critical Transnational Research Beyond Anglocentrism, *Harvard Kennedy School Misinformation Review,* Volume 3(2). [accessed 27 September 2023].
[1805] Carnegie Endowment for International Peace (Goodwin, C. and Jackson, D.), 2022.

16.51 Gendered disinformation, often targeted towards women, has been defined by Demos as information activity (including creating, sharing and disseminating content) that attacks or undermines people based on their gender, or weaponises gendered narratives to promote political, social or economic objectives.[1806]

16.52 Research has found that political leaders in Russia, the Philippines, Hungary and Turkey have used gendered disinformation campaigns to attack women in politics.[1807] [1808] Gendered disinformation has been used by foreign interference perpetrators to target high-profile women, especially politicians, journalists and women's rights activists, often to discredit, target or silence them.[1809]

16.53 When women are targeted by disinformation or foreign influence operations, they are often targeted in ways that are specific to their gender, or in ways that rely on gendered assumptions, narratives or expectations. This includes the use of tactics that specifically draw on gendered dynamics; for example, when female politicians are targeted with death, rape or sexual assault threats, or deepfake pornography of themselves.

16.54 Some evidence suggests that diaspora groups may be disproportionately at risk of harm from foreign interference operations. In the UK, Freedom House has reported that members of exiled and diaspora groups who might serve as subjects or sources for British media reporting on the People's Republic of China have faced online trolling and Chinese-state-led intimidation in the UK, including the targeting of prominent Hong Kong politician and activist Nathan Law.[1810] Another example includes the former Foreign, Commonwealth and Development Office Minister's mention of reports that members of the Uighur diaspora in the UK were being harassed and intimidated by the Chinese authorities to silence and force them to return to the People's Republic of China, or to share information on other Uighurs.[1811]

16.55 A report into the malign influence activities of Russian-state-linked influence operation Doppelganger found that content created in a campaign run by the group is critical of LGBTQ+ rights and inclusivity efforts in the US, and is likely intended to fuel hostile rhetoric and amplify anti-LGBTQ+ sentiments in the US. As a result, LGBTQ+ individuals may have an increased risk of harm from foreign interference operations that deliberately seek to amplify anti-LGBTQ+ narratives.[1812] Research from the Institute for Strategic Dialogue has found that recorded hate crimes against LGBTQ+ people have increased in multiple

---

[1806] Demos (Judson, E., Atay, A., Krasodomski-Jones, A., Lasko-Skinner, R. and Smith, J.), 2020. Engendering Hate: The Contours of State-Aligned Gendered Disinformation Online. [accessed 6 March 2023].
[1807] Di Meco, L. and Wilfore, K., 2021. Gendered disinformation is a national security problem, *Brookings Institute*, 8 March. [accessed 6 March 2023].
[1808] For example, Ukrainian MP Svitlana Zalishchuk was targeted after her speech to the United Nations on the effects of the war with Russia. The campaign included a screenshot of a falsified tweet claiming she had promised to run through Kyiv naked if the Ukrainian army lost an important battle, accompanied with fake images of her naked. Zalishchuk has suggested that the campaign originated in Russia, as it began during a period of high tension between Russia and Ukraine, and the fake claims and images first appeared on pro-Kremlin platforms. Source: HM Government Stabilisation Unit, 2020. Quick-read guide: gender and countering disinformation. [accessed 6 March 2023].
[1809] HM Government Stabilisation Unit, 2020. Quick-read guide: gender and countering disinformation. [accessed 6 March 2023].
[1810] Freedom House (Datt, A. and Dunning, S.), 2022. Beijing's Global Media Influence 2022. [accessed 17 February 2023].
[1811] Hope, C., 2021. Exclusive: Uighurs harassed and abused by Beijing in UK, minister admits, *The Telegraph*, 13 March. [accessed 17 February 2023].
[1812] Recorded Future, 2023. Obfuscation and AI Content in the Russian Influence Network "Doppelganger" Signals Evolving Tactics. [accessed 16 May 2024].

jurisdictions, including in the UK, alongside increased online anti-LGBTQ+ conversations[1813], suggesting that foreign interference operations advancing anti-LGBTQ+ narratives may have the potential to present a risk of offline harm to LGBTQ+ individuals.

16.56 Some evidence suggests that different religious groups may be at an increased risk of harm from foreign interference operations. A report into anti-Hindu disinformation by the Network Contagion Research Institute and Rutgers University found that Iranian state-sponsored trolls disseminated anti-Hindu stereotypes on Twitter whilst self-reporting their locations to be in Pakistan and occasionally India. This included running a campaign on Twitter during the aftermath of the 2020 Delhi riots, alleging that Hindus were committing genocide against Muslims during the unrest.[1814]

16.57 There is some evidence that individuals with intersectional identities (overlapping or intersecting social identities, such as women of ethnic minority backgrounds) are at an increased risk of harm. A study into the gendered abuse and disinformation directed at thirteen female politicians in English-speaking countries found that women of colour were subjected to compounded and intersectional narratives that targeted both their gender and their race or ethnicity.[1815]

Risk factors: Functionalities and recommender systems

User identification

User profiles

16.58 The evidence we assessed suggests that the ability to target specific sub-groups on a service can be exploited by perpetrators of foreign influence operations, particularly where personal information is visible on user profiles. Diaspora groups may be targeted for foreign influence operations (see Risk factors: user base section above). It may be possible that the display of profile information which would enable other users to identify members of these groups (for example, information about language spoken or home town) could also increase the risks of harm to these diaspora groups.

16.59 There is also some evidence suggesting the targeting of users by political ideology on specific social media services with divisive narratives.[1816]

Fake user profiles

16.60 The evidence we have assessed suggests that fake user profiles can be exploited by perpetrators of foreign interference operations. A newspaper investigation in February 2023 found a sophisticated unit of disinformation operatives claiming to have manipulated more than 30 elections worldwide. [1817] The unit claims to use a variety of tactics, including automated disinformation on a range of social media services by creating fake accounts.

[1813] Institute for Strategic Dialogue (Squirrel, T. and Davey, J.), 2023. A Year of Hate: Understanding Threats and Harassment Targeting Drag Shows and the LGBTQ+ Community. [accessed 16 May 2024].
[1814] Network Contagion Research Institute, 2022. Quantitative Methods for Investigating Anti-Hindu Disinformation. [accessed 16 May 2024].
[1815] Wilson Center (Jankowicz, N., Hunchak, J., Pavliuc, A., Davies, C., Pierson, S., and Kaufmann, Z.), 2021. Malign Creativity: How Gender, Sex and Lies are Weaponised Against Women Online. [accessed 16 May 2024].
[1816] Graphika, 2020. Step into my Parler. [accessed 22 September 2023].
[1817] Kirchgaessner, S., Ganguly, M., Pegg, D., Cadwalladr, C. and Burke, J., 2023. Revealed: the hacking and disinformation team meddling in elections, *The Guardian*, 15 February. [accessed 15 February 2023].

According to those reports, the operatives also sold a product which enables the simple creation of fake accounts on several U2U services, which they claim to have sold to unnamed intelligence agencies, political parties and corporate clients.[1818] Journalists also found evidence that campaigns using this product had previously targeted the UK, as well as other countries.[1819]

16.61 In another example, TikTok reported a covert influence operation running a network of 1,686 fake accounts, operating from Russia and targeting Germany, Italy, the UK and other European countries. The network used localised, fake accounts and used speech synthesis to share content in German, Italian and English to amplify pro-Russian viewpoints and target discourses about the Russian invasion of Ukraine.[1820]

16.62 Fake user profiles can also be used to hide identity and impersonate authoritative and high-profile sources. A report by EU Parliament Rapporteur Sandra Kalniete on foreign interference in the EU notes that fake personas and identities are used within these foreign influence efforts.[1821]

16.63 Meta's January 2021 *Coordinated Inauthentic Behaviour Report* detailed an influence operation on Facebook that impersonated legitimate think-tanks and media organisations in Israel and the UK, and shared content through them.[1822] Additionally, in 2022, Meta reported on a Russian-origin network that mimicked the exact layout and spoofed web addresses of mainstream European media outlets like Der Spiegel, the Guardian and Italian news agency ANSA.[1823] The content created by these sites was amplified on social media by many fake accounts, which often claimed to work for organisations like Netflix.

16.64 Recorded Future, a threat intelligence organisation, found over 800 automated social media accounts were used by Russian state-linked actors to share links to 'inauthentic' articles posted by inauthentic media outlets impersonating multiple, reputable Ukrainian media organisations. This set of tactics, techniques and procedures have been used several times by a Russia-linked information operation labelled 'Doppelgänger' and this series of actions have also been linked to this ongoing operation. These articles spread narratives undermining Ukraine's military strength, political stability, and international relationships with Western allies.[1824]

16.65 Some emerging evidence also shows the increasing integration of generative AI in these campaigns. Recorded Future found that the Russia-linked 'Doppelgänger' information operation created an inauthentic media outlet to target US audiences with many articles published on the site flagged as either partially or nearly wholly written by AI. These articles are shared by Doppelgänger's network of inauthentic social media accounts, and target US audiences with content focused on US election cycles, polling, and political campaigning.

[1818] Kirchgaessner, S., Ganguly, M., Pegg, D., Cadwalladr, C. and Burke, J., 2023.
[1819] Ganguly, M., 2023. 'Aims': the software for hire that can control 30,000 fake online profiles, *The Guardian*, 15 February. [accessed 15 February 2023].
[1820] TikTok, 2022. Community Guidelines Enforcement Report. accessed 15 February 2023].
[1821] REPORT European Parliament (Kalniete, S.), 2022. REPORT on foreign interference in all democratic processes in the European Union, including disinformation. [accessed 27 September 2023].
[1822] Facebook, 2021. January 2021 Coordinated Inauthentic Behaviour Report. [accessed 28 June 2023].
[1823] Meta, 2022. Removing Coordinated Inauthentic Behavior From People's Republic of China and Russia. [accessed 27 September 2023].
[1824] Recorded Future, 2023. Obfuscation and AI Content in the Russian Influence Network "Doppelganger" Signals Evolving Tactics. [accessed 16 May 2024].

Utilising a similar methodology, Recorded Future found another Russian state actor-linked campaign, 'CopyCop', targeting audiences in the US, UK and France with content plagiarised from mainstream media outlets via generative AI, and weaponised by introducing partisan bias to coverage of domestic news in targeted countries, alongside pro-Russian coverage of the war in Ukraine and coverage of the Israel-Hamas conflict that is critical of Israeli military operations in Gaza.[1825]

16.66 Research has shown that multiple, often fake, user profiles can be created rapidly in foreign influence campaigns. A study by the Oxford Internet Institute found a network of 62 Twitter accounts, active between June 2020 and January 2021, that amplified and engaged with the Twitter accounts of diplomats from the People's Republic of China in the UK. Almost a third of the accounts were created within minutes of each other, with many of them sitting dormant for months at a time and then activating in unison at specific moments, and many impersonated UK citizens (for example, account names containing 'UK').[1826]

Anonymous user profiles

16.67 The evidence we have assessed suggests that the ability to create anonymous user profiles and to post anonymously can be exploited by perpetrators of foreign influence operations. Some discussion forums allow unregistered users to post content anonymously without creating an account, which can then be used in foreign influence operations as well as to spread disinformation on the services.[1827] For example, it is alleged that anonymous users of a service were involved in foreign influence operations during the 2017 French Presidential election.[1828]

16.68 Research has found that multiple foreign influence operations conducted by the People's Republic of China have manipulated Twitter's policies on anonymous accounts to raise their diplomats' profiles and amplify their messaging on the platform.[1829]

User networking

User connections

16.69 The evidence we have assessed suggests that the use of coordinated networks on social media accounts can be exploited by the perpetrators of foreign influence operations. The ability to connect with users can allow such networks to be built. These coordinated networks can be used to amplify content (as in the following example) to wider audiences and can be used to spread narratives across social media platforms through the simultaneous posting of similar or identical content.

16.70 For example, in the previously referenced Oxford Internet Institute study, the authors identified a network of 62 Twitter accounts impersonating members of the public in the UK

[1825] Recorded Future, 2024. Russia-linked CopyCop Uses LLMs to Weaponise Influence Content at Scale. [accessed 16 May 2024].
[1826] Schliebs, M., Bailey, H., Bright, J. and Howard, P. N., 2021. People's Republic of China's Inauthentic UK Twitter. Diplomacy: A Coordinated Network Amplifying PRC Diplomats, *Programme on Democracy & Technology*. [accessed 27 September 2023].
[1827] In such cases, the identity of users may also be unknown to services.
[1828] Glaser, A., 2017. Macron's French presidential campaign has been hacked less than 48 hours before the election, *Vox,* 6 May. [accessed 22 September 2023].
[1829] Schliebs, M., Bailey, H., Bright, J. and Howard, P. N., 2021. People's Republic of China's Inauthentic UK Twitter Diplomacy: A Coordinated Network Amplifying PRC Diplomats, Programme on Democracy & Technology. [accessed 27 September 2023].

that amplified and engaged with UK-based diplomats from the People's Republic of China. Many of the accounts in this network followed, and focused on, these diplomats, with the sole aim of raising their profile in the UK.[1830]

User groups

16.71 The evidence we have assessed suggests that the creation and use of groups can be exploited by perpetrators of foreign influence campaigns.

16.72 It is well evidenced that services with group functionalities can instigate users to conduct offline activity,[1831] and foreign influence operations have been found to use this to their advantage. For example, in 2016, Russia's Internet Research Agency used Facebook groups to organise a protest and counter-protest in Houston, Texas to create division and tension within the community.[1832]

16.73 In addition, in January 2022, Meta reported that it had removed a small network of Facebook accounts originating in St. Petersburg, Russia, which targeted Nigeria, Cameroon, Gambia, Zimbabwe and Congo. One of the tactics the network used was trying to solicit freelance help to write articles about Syria through Arabic-language journalist groups.[1833] In February 2022 Meta removed a small network: 27 Facebook accounts, two pages, three groups and four Instagram accounts, which had originated in Russia and targeted people in Ukraine, promoting claims that the West had betrayed Ukraine and that Ukraine was a failed state.[1834]

User tagging

16.74 The evidence suggests that the ability to tag users can be exploited by perpetrators of foreign interference campaigns.

16.75 During the UK 2019 General Election, leaked documents detailing trade talks between the US and the UK were posted on Reddit, Twitter and across several websites by accounts that suggested links to the Russian influence operation, Secondary Infektion;[1835] Reddit attributed a network of 61 accounts sharing these documents to Russia.[1836] A Twitter account was used to promote links to the leaked documents by tagging opposition politicians and prominent journalists.[1837]

User communication

Direct messaging, group messaging, encrypted messaging

16.76 The creation and amplification of disinformation content is a primary component of many foreign influence operations. Think-tank and campaign group First Draft's 'Trumpet of Amplification' highlights how disinformation actors have used anonymous online spaces to

---

[1830] Schliebs, M., Bailey, H., Bright, J. and Howard, P. N., 2021.
[1831] Thiel, D. and McCain, M. 2022. Gabufacturing Dissent: An in-depth analysis of Gab, Stanford Cyber Policy Review. [accessed 27 September 2023].
[1832] Franceschi-Bicchierai, L., 2017. Russian Facebook Trolls Got Two Groups of People to Protest Each Other in Texas, *Vice*, 1 November. [accessed 17 February 2023].
[1833] Meta, 2022. January 2022 Coordinated Inauthentic Behaviour Report. [accessed 17 February 2023].
[1834] Meta (Nimmo, B., Agranovich, D. and Gleicher, N.), 2022. Adversarial Threat Report. [accessed 17 February 2023].
[1835] Graphika (Nimmo, B.), 2019. UK Trade Leaks. [accessed 27 September 2023].
[1836] Wendling, M., 2019. General election 2019: Reddit says UK-US trade talks document leak 'linked to Russia, *BBC News*, 7 December. [accessed 27 September 2023].
[1837] Graphika (Nimmo, B.), 2019.

create rumours and place fabricated content, spreading from these encrypted spaces to closed and semi-closed networks, to conspiracy communities, then mainstream social media, to finally end up being reported on in the mainstream media.[1838] Further evidence demonstrated that encrypted applications lack the conventional fact-checking and content moderation that is offered on other services, thereby offering a unique opportunity to those wishing to easily spread disinformation.[1839]

Reacting to content

16.77 Some evidence suggests that the ability to engage with another user's content could be exploited by perpetrators of foreign influence campaigns. For example, users can purchase 'likes' to enable fake profiles to promote selected content and inflate its popularity on social media services.[1840]

Posting content

16.78 **Multimodal disinformation** is one of the most reported tactics used in foreign interference operations. It combines image and text formats to create false or misleading content. These can include (a) *de-contextualisation:* when real images or videos are paired with false, manipulated or misleading text; and (b) *multimodal doctoring:* when content is fabricated by pairing manipulated images or videos with false, misleading or manipulated content.[1841] Other types of multimodal disinformation are outlined in the 'content editing' section.

16.79 Examples of these forms of multimodal disinformation include memes (photos paired with small snippets of text) which were frequently repurposed to target US social media users during the 2016 US Presidential elections. The US Congressional investigation into the Internet Research Agency's activities received over 100,000 memes from Instagram and 67,000 memes from Facebook.[1842]

16.80 Some evidence shows the use of Generative AI to create content, which when posted on social media services, messaging services, video-sharing platforms and other services, have been used in foreign interference operations. This content can take various forms, including audio, video, image and text-based forms.

16.81 The evidence outlined in this section does not focus on confirmed examples of foreign interference. Rather, the evidence shows how these technologies could be utilised to create and disseminate content by perpetrators of a potential foreign interference offence.

16.82 For example, two days before the Slovak Republic's election in 2023, deepfake audio recordings allegedly featuring conversations between a journalist and Michal Šimečka, a

---

[1838] Wardle, C., 2018: 5 Lessons for Reporting in an Age of Disinformation, *First Draft News*, 27 December. [accessed 21 September 2023].
[1839] Gurksy, J., Riedl, M. J. and Woolley, S., 2021. The Disinformation Threat to Diaspora Communities in Encrypted Chat Apps, *Brookings Institute,* 19 March. [accessed 27 September 2023].
[1840] Koval, I., 2021. "How social media is manipulated — and how Russia is involved", *DW*, 14 April, [accessed 22 June 2023].
[1841] Hameleers, M., Powell, T. E., Van Der Meer, T. G.L.A. and Bos, L., 2020. "A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated on Social Media", Political Communication, 37(2), p.281-301. [accessed 27 September 2023].
[1842] DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. and Johnson, B., 2019. The Tactics and Tropes of the Internet Research Agency, New Knowledge. [accessed 27 September 2023].

leading liberal politician, about vote-rigging and other controversial issues, appeared on social media and video-sharing platforms.[1843]

16.83 British politicians have also been targeted by deepfake audio and video content, with an audio clip alleging to depict the then-Leader of the Opposition Sir Keir Starmer swearing at a member of his staff posted to several social media services in October 2023. Fact-checking organisation Full Fact found no evidence that the clip was genuine.[1844] [1845] London Mayor Sadiq Khan has also been targeted by deepfake audio, with a scripted, AI-generated replica of his voice used to suggest had made inflammatory remarks ahead of Armistice Day in 2023 and shared widely on social media. While neither example has been linked to a state-actor-associated perpetrator, they illustrate the ease with which a foreign interference perpetrator may create a realistic-seeming fake audio of a politician for nefarious purposes.[1846]

16.84 A report by media and research consultancy Fenimore Harper Communications found over 143 deepfake video advertisements impersonating former Prime Minister Rishi Sunak, alongside a series of journalists, on a popular social media service between 8 December 2023 and 8 January 2024, promoting various false investment schemes.[1847]

16.85 We have also found evidence of Generative AI models being used to create text-based content as part of foreign interference campaigns. Recorded Future found that the Russia-linked 'Doppelgänger' information operation created an inauthentic media outlet to target US audiences with many articles published on the site flagged as either partially or nearly wholly written by AI. These articles are shared by Doppelgänger's network of inauthentic social media accounts, and target US audiences with text-based content focused on US election cycles, polling, and political campaigning.[1848]

16.86 Utilising a similar methodology, Recorded Future found another Russian state actor-linked campaign, CopyCop, targeting audiences in the US, UK and France with text-based content plagiarised from mainstream media outlets via generative AI, and weaponised by introducing partisan bias to coverage of domestic news in targeted countries, alongside pro-Russian coverage of the war in Ukraine and coverage of the Israel-Hamas conflict that is critical of Israeli military operations in Gaza.[1849] Further examples of Generative AI content used in foreign interference campaigns can be found in the *Editing content* section.

[1843] Council for Media Services, 2024. Monitoring of platform functionalities in relation to the 2023 Elections to the National Council of the Slovak Republic. [accessed 16 May 2024]
[1844] Full Fact, 2023. No evidence that audio clip of Keir Starmer supposedly swearing at staff is genuine. [accessed 16 May 2024].
[1845] Resemble.ai, 2023. Political Deepfake: Keir Starmer. [accessed 16 May 2024]
[1846] Spring, A., 2024. Sadiq Khan says fake AI audio of him nearly led to serious disorder, BBC News, 13 February. [accessed 26 September 2024].
[1847] Fenimore Harper Communications (Beard, M.), 2024. Over 100 Deep-Faked Rishi Sunak Ads Found on Meta's Advertising Platform. [accessed 16 May 2024].
[1848] Recorded Future, 2023. Obfuscation and AI Content in the Russian Influence Network "Doppelganger" Signals Evolving Tactics. [accessed 16 May 2024].
[1849] Recorded Future, 2024. Russia-linked CopyCop Uses LLMs to Weaponise Influence Content at Scale. [accessed 16 May 2024].

Re-posting or forwarding content

16.87 Frequently forwarded and mass-forwarded messages, sent directly from a trusted set of contacts on an encrypted private messaging service to more open services,[1850] have been a primary mechanism for the spread of disinformation, including foreign influence operation content. This prompted, for example, WhatsApp to introduce forwarding limits in 2020 following the onset of the COVID-19 pandemic.[1851]

16.88 Posting or retweeting content can create an artificial engagement that could lead to manipulation. For example, a People's Republic of China-linked network of accounts, studied by the Oxford Internet Institute, accounted for 44% of retweets received by the Chinese Ambassador to the UK, and 30% of retweets received by the Chinese Embassy to the UK between June 2020 and January 2021. Amplification networks, like this example on Twitter, can manipulate user recommendations, and therefore artificially amplify content.[1852]

16.89 The 'doxxing' of individuals is another tactic used within foreign influence operations. Doxxing refers to the malicious sharing of individuals' private information, including email addresses, physical addresses, phone numbers and social security information online, without the individual's permission. This information is often obtained illicitly – for example, via a hack and leak operation. An example of this is the social media activity 'Project Nemesis', a Russian hacking group which doxxed members of the Ukrainian military, secret services, volunteers and international trainers supporting Ukraine in resisting Russia's invasion. The level of state involvement in this operation is unclear.[1853]

Content exploring

Hyperlinking

16.90 Many foreign influence operations are run concurrently across different platforms and hyperlinking is an important tool for sharing content across or between platforms. For instance, during the Internet Research Agency campaign targeting the 2016 US Presidential election, Russian state media outlets including Sputnik, RT and Ria Novosti embedded tweets from different Internet Research Agency-linked accounts into their reporting[1854]; hyperlinks to inauthentic articles were shared by inauthentic social media accounts targeting US audiences with text-based content focused on US election cycles, polling, and political campaigning as part of the Russia-linked 'Doppelganger' information operation. [1855]

16.91 The evidence we have assessed suggests that sharing hyperlinks can be exploited by perpetrators of the foreign influence offence in several ways, including by highlighting

---

[1850] Gurksy, J., Riedl, M. J. and Woolley, S., 2021. The Disinformation Threat to Diaspora Communities in Encrypted Chat Apps, *Brookings Institute,* 19 March. [accessed 27 September 2023].
[1851] WhatsApp, n.d. About forwarding limits | WhatsApp Help Center. [accessed 27 September 2023].
[1852] Schliebs, M., Bailey, H., Bright, J. and Howard, P. N., 2021. People's Republic of China's Inauthentic UK Twitter Diplomacy: A Coordinated Network Amplifying PRC Diplomats, Programme on Democracy & Technology. [accessed 27 September 2023].
[79] Institute for Strategic Dialogue (Thomas, E.), 2022. Project Nemesis, Doxxing and the New Frontier of Informational Warfare. [accessed 27 September 2023].
[1854] DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. and Johnson, B., 2019. The Tactics and Tropes of the Internet Research Agency, New Knowledge. [accessed 27 September 2023].
[1855] Recorded Future, 2023. Obfuscation and AI Content in the Russian Influence Network "Doppelganger" Signals Evolving Tactics. [accessed 16 May 2024]

content gained illicitly through hack and leak operations, and by amplifying doxxing campaigns[1856].

16.92 On Gab, links to YouTube videos are the most-posted destination domain on the platform, with some studies suggesting that several views of extremist and disinformation content on YouTube are driven by external referral sites like Gab. Other commonly-linked-to sites include Facebook, Twitter and closed Telegram groups.[1857]

16.93 'Hack and leak operations' are another tactic used in foreign influence operations to disseminate and draw attention to documents online. In December 2019, Reddit attributed a network of 61 accounts, sharing hyperlinks leading to leaked documents detailing US toUK trade documents, to Russia.[1858] A similar operation is reported to have been used to target and dox allegedly British-influenced individuals in Russia on social media.[1859]

Content editing

Editing visual media

16.94 Cheap-fakes[1860] and deepfakes[1861] have been a primary concern for information operations researchers since they came to prominence in 2017.[1862] This is due to their ease of deployment and varied potential uses in foreign interference operations. Two common tactics to create these include (a) *reframing,* where videos are cropped or decontextualised to make certain aspects more prominent in pursuit of a specific agenda, and (b) *visual doctoring,* where images or videos are manipulated to present a different reality to that in their non-edited form.[1863] These are different types of **multimodal** disinformation (see 'posting content' for more).

16.95 Examples include a falsified BBC report presenting a fake story of a nuclear escalation between Russia and NATO which began circulating on WhatsApp. The incident has never been definitively attributed.[1864] Another example is a low-quality deepfake of Ukrainian President Volodymyr Zelenskyy talking about surrendering to Russia on social media

---

[1856] Institute for Strategic Dialogue (Thomas, E.), 2022. Project Nemesis, Doxxing and the New Frontier of Informational Warfare. [accessed 27 September 2023].
[1857] Thiel, D. and McCain, M., 2022. Gabufacuturing Dissent: An in-depth analysis of Gab, Stanford Cyber Policy Review. [accessed 27 September 2023].
[1858] Wendling, M., 2019. General election 2019: Reddit says UK-US trade talks document leak 'linked to Russia', *BBC*, 7 December. [accessed 27 September 2023].
[1859] Institute for Strategic Dialogue, 2022. Tales From the Underside: A Kremlin-Approved Hack, Leak & Doxxing Operation. [accessed 27 September 2023].
[1860] Cheap-fakes are videos that use conventional video editing techniques like speeding, slowing, cutting, restaging or re-contextualising video footage
[1861] Deepfakes are a specific type of media that involves the use of AI algorithms, particularly generative AI models, to modify videos, images or audio to create realistic synthetic content. This is often done by superimposing the face of a person onto the body of another person in a video or image as well as voice manipulation with lip syncing. Deepfakes are shared as user generated content on user-to-user services but could also potentially be created using functionalities present on user-to-user services. Deepfake technology is currently used to create content that can be harmful; however, we acknowledge that it may also have positive use cases.
[1862] Donovan, J. and Paris, B., 2019. Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence, *Data & Society.* [accessed 27 September 2023].
[1863] Hameleers, M., Powell, T. E., Van Der Meer, T. G.L.A. and Bos, L., 2020. A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated on Social Media, *Political Communication*, 37(2), pp.281-301. [accessed 27 September 2023].
[1864] Donovan, J. and Paris, B., 2019. Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence, *Data & Society.* [accessed 27 September 2023].

services.[1865] Although the video has not been specifically attributed to a state actor, it is widely believed that it was generated by Russia.

16.96 There is also an example of the first confirmed case of a state-aligned information operation using deepfakes, viewed less than 300 times. It involved AI-generated footage of fictitious people being promoted by Spamouflage, a pro-Chinese influence operation.[1866]

16.97 In June 2023, the Russian Embassy in the UK posted a deepfake video on sanctions on Russia of US President Joe Biden, UK Prime Minister Rishi Sunak, European Commission President Ursula von der Leyen and French President Emmanuel Macron.[1867]

Editing posted content

16.98 The evidence we have assessed suggests that retroactively editing posts on social media services may be exploited by perpetrators of foreign influence operations. For example, disinformation and misinformation researchers raised concerns regarding Twitter's plans to roll out an 'edit' button, as a feature that could be exploited by malicious actors, including hostile state actors and those involved in foreign influence operations.[1868]

16.99 There is also evidence of risk in editing public content, such as on information-sharing websites services like Wikipedia. A study by ISD and CASM Technology found several editing-based tactics used by perpetrators on Wikipedia[1869] which could be vulnerable to influence operations, including foreign influence operations. Other practices include undisclosed paid editing, adversarial editing and state editing.

16.100 Undisclosed paid editing can be carried out by 'reputation management' or 'reputation protection' providers, which violates Wikipedia's policies if it misuses an editor's power of office or is undisclosed.[1870]

16.101 Adversarial editing refers to the concerted and coordinated attempts to edit pages for ideological or political reasons, including to celebrate or promote a specific group, or to 'get the truth out' about an event, conflict or controversy. It is characterised by continuous revisions and counter-revisions across one page or a group of pages over time. ISD and CASM Technology found an example of 89 editors who made changes to an article on the Russo-Ukraine war which exhibited narratives consistent with those spread in Russian state-sponsored information warfare.

16.102 State editing can happen when individuals linked to a foreign state edit information to promote the aims and goals of the foreign policy. An example of state editing was seen in the banning of seven editors linked to a group from the People's Republic of People's Republic of China.[1871]

[1865] Wakefield, J., 2022. Deepfake presidents used in Russia-Ukraine war, *BBC News*, 18 March. [accessed 6 March 2023].
[1866] Graphika, 2023. Deepfake It Till You Make It. [accessed 17 February 2023].
[1867] Thurston, J, 2023. Russia deepfake video mocks Rishi Sunak and Joe Biden, *The Times,* 15 June. [accessed 27 September 2023].
[1868] Robson, K., 2022. Will Twitter's edit button help spread more fake news?, *Verdic*t, 2 September 2022. [accessed 27 September 2023].
[1869] Institute for Strategic Dialogue and CASM Technology (Miller, C., Smith, M., Marsh, O., Balint, K., Inskip, C, and Visser, F.), 2022. Information Warfare and Wikipedia. [accessed 27 September 2023].
[1870] Institute for Strategic Dialogue and CASM Technology, 2022.
[1871] Institute for Strategic Dialogue and CASM Technology (Miller, C., Smith, M., Marsh, O., Balint, K., Inskip, C, and Visser, F.), 2022. Information Warfare and Wikipedia. [accessed 27 September 2023].

Editing usernames

16.103 Evidence suggests that the ability to change a username, handle, or other information presented on a user profile can be exploited by perpetrators of foreign interference operations. Researchers at the City University of London found that 26,538 Twitter accounts suddenly changed their usernames after the EU referendum in 2016, and 5% of all Twitter accounts that had tweeted about the referendum were either deleted or renamed.[1872] This changing of account names can be used by perpetrators to quickly repurpose accounts from one influence operation to another. It also enables perpetrators to easily change the focus of an account if it is not performing as well as they would like, or if they want to switch focus to a different topic. There are examples of the Internet Research Agency renaming and rebranding some of its Instagram accounts during its operation targeting the 2016 US Presidential election.[1873]

Recommender systems

Content recommender systems

16.104 In addition to personalisation, content recommender systems are commonly designed to suggest content that might be trending or popular (measured by number of likes, shares, or comments). Such systems are understood to learn about popular and trending content through the volume of user feedback; this normally includes explicit feedback (active engagement such as reactions, posts and comments) and implicit feedback (viewing the content many times but not necessarily engaging with it).[1874] This fundamental characteristic of recommender system design leaves services vulnerable to manipulation by third parties, particularly if their design is simple (for example, if all content is ranked in the same way and all types of engagement are registered as positive feedback on all types of content). We consider that content recommender systems may be manipulated by perpetrators of foreign influence operations.

Risk factors: Business model and commercial profile

Revenue models

16.105 Evidence suggests that services which raise income through advertising may be exploited by bad actors who can use advertisements to spread foreign interference content. A report on the political ad policy for an online U2U service recognised this potential risk,[1875] saying that *"scrutiny of major online advertising platforms intensified due to foreign interference in the 2016 U.S. elections as well as broader concerns on disinformation, voter suppression, and inauthentic behaviour"*.[1876] The report added that if users are unaware of the political intent

---

[1872] Bastos, M. T. and Mercea, D, 2017. The Brexit Botnet and User-Generated Hyperpartisan News, *Social Science Computer Review*, 37(1). [accessed 27 September 2023].

[1873] DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. and Johnson, B., 2019. The Tactics and Tropes of the Internet Research Agency, New Knowledge. [accessed 27 September 2023].

[1874] Perpetrators may take advantage of the design of engagement focused systems by acting in a coordinated fashion to generate high volumes of explicit feedback to artificially inflate the dissemination of specific posts, for example, by using multiple accounts to upload and share the same content many times. By artificially inflating engagement, content recommender systems could then be more likely to promote this content to users.

[1875] The report suggested that *"online political advertising is a powerful tool for enabling engagement in the political process but that with this power comes the risk of abuse that can harm the integrity of the democratic process."*

[1876] Le Pochat, V., Edelson, L., Van Goethem, T., Joosen, W., McCoy, D. and Lauinger, T., 2022. An audit of Facebook's Political Ad Policy Enforcement. [accessed 27 September 2023].

behind the advert, the adverts can be more effective in their malicious intent. Hence, services offering advertising, without effective moderation policies to identify and label adverts that seek to influence public political FI opinion as 'political', are more able to be used by malicious advertisers and thereby weaken the integrity of the online political ad ecosystem.[1877]

16.106 Some evidence suggests that the ability to purchase adverts on a service can be exploited by perpetrators of foreign influence campaigns. During its 2016 operation targeting the US Presidential election, a study found that the Internet Research Agency created 1,852 adverts using Facebook's interest-based targeting functions, mostly focusing on African-American interests and communities. Some of these adverts used geographical targeting, some targeted users via gender and others used more specific categories, with one notable advert targeting users with the job title 'Coal Miner'.[1878]

16.107 The same study recognises how ad targeting[1879] may affect the risk of foreign interference in advertising revenue models. The user data that online advertising uses may make foreign interference operations more effective in targeting specific segments of the population with their adverts.

16.108 A report suggests that social media services can generate valuable revenues from advertisers whose intention is to manipulate behaviour in a coordinated manner, and crucially, without accurately identifying the source behind the advertisement.[1880]

[1877] Le Pochat, V., Edelson, L., Van Goethem, T., Joosen, W., McCoy, D. and Lauinger, T., 2022.
[1878] DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. and Johnson, B., 2019. The Tactics and Tropes of the Internet Research Agency, New Knowledge. [accessed 27 September 2023].
[1879] Some services enable advertisers to purchase adverts and target them at specific users based on information that the users have provided to services, and data that the services have gathered on users' interests and behaviour.
[1880] Colliver, C., King, J. and Maharasingam-Shah, E., 2020. Hoodwinked: Coordinated Inauthentic Behaviour on Facebook. [accessed 27 September 2023].

17. Animal cruelty

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for animal cruelty offences: How harms manifest online, and risk factors

This section summarises the risks of harm to individuals from the animal cruelty offence.

The existence of online activities that encourage, assist or commit acts of animal cruelty may result in content being made available which may distress a user, or cause them to engage in harmful or illegal behaviours and activities themselves. This section covers several factors which could be associated with the offence. Of these, we consider the following to be key and have included them in the Risk Profiles (see Section 7).

Service type risk factors:

As with almost all kinds of illegal harm, **social media services** are a prominent risk factor for this offence, since content depicting cruelty to animals (which may in itself encourage, assist or conspire to further animal cruelty) shared on these services can receive wide reach.

Our evidence also points to **messaging services** being a risk factor, in that they allow perpetrators to form a community and to discuss ideas for, acts of cruelty. They may also assist in the production and publication of animal cruelty content or share it via the messaging services.

Functionalities risk factors:

Services with the ability to **post images or videos** (which may be social media services) pose a significant risk of this offence, particularly where they can encourage or facilitate further acts of animal cruelty. **Commenting on content** can also enable people to encourage, assist or conspire to commit further acts of animal cruelty.

Our evidence also points to services **where users can form user groups or send group messages** being a prominent functionality that is a risk factor, for the same reason that we believe messaging services to be a risk factor – it may allow perpetrators to come together to discuss and encourage, facilitate or commit acts of animal cruelty. This may be especially the case for closed and private groups.

Introduction

17.1 This section summarises our assessment of the risks of harm to individuals presented by:

- content on U2U and search services that may amount to the animal cruelty offences listed under 'Relevant offences' below; and

- the use of U2U services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

17.2 We set out the characteristics of U2U services and, so far as possible, search services, that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

Relevant offences

17.3 The Act requires Ofcom to consider the risk of harm connected with the priority offences as set out in the Act. This section considers the risk of harm connected with the offence that concerns the unnecessary suffering of animals (section 4(1) of the Animal Welfare Act 2006), referred to as the animal cruelty offence. The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of priority offences.

17.4 We recognise that most acts of cruelty occur offline. The publication *online* of content relating to or depicting these offline acts does not in itself cause the animal unnecessary suffering (or further suffering) and therefore cannot constitute an offence under the Animal Welfare Act. Where content depicts past instances of animal cruelty, providers should consider the evidence provided on the non-priority offence under s.127(1) of the Communications Act in the Register of Risks for obscene content showing torture of humans and animals, and the Illegal Content Judgements Guidance.

17.5 However, where a user publishes content showing, describing or discussing cruelty to animals in order to encourage, assist or conspire to commit acts of animal cruelty, this would create priority illegal content under the Act. This is because encouragement, assistance and conspiracy to commit a priority offence are each priority offences in their own right. In our view, a livestream of animal cruelty being carried out, which users choose to watch knowing what they will see, can be characterised as a conspiracy to commit the animal cruelty offence and is likely to amount to priority illegal content.

17.6 For more details on how providers can assess whether content amounts to the animal cruelty offence, see our Illegal Content Judgements Guidance.

Use of the service for commission or facilitation

17.7 A user may also use online services to facilitate or commit acts of animal cruelty. In some cases, it may be that the ability to disseminate animal cruelty content online is part of the motivation for the animal cruelty offences being committed.

17.8 Online services may enable networks of bad actors to form, share ideas, view and engage with animal cruelty content, and arrange for future activities. They may also provide a medium on which these acts may be advertised or where content made about them may be subsequently sold and shared online. Content published for this purpose may then in itself potentially encourage *further* illegal acts or be part of a conspiracy to do so.

Other offences

17.9 Acts of animal cruelty have demonstrable links to other harmful acts which may manifest online, such as extreme pornography (bestiality) and child abuse.

- **Extreme pornography:** the depiction of the sexual abuse of animals is an example of the animal cruelty and torture content which could manifest online (and is discussed in this section). With regard to bestiality, Ofcom recognises that possession of an image depicting the sexual abuse of an animal is an illegal offence of possession of extreme pornography and a primary priority harm under the Online Safety Act 2023. The extreme pornography offence is explored in the 'Extreme pornography' chapter.
- **Child abuse and grooming**: patterns of child abuse can be escalated, with the perpetrator inciting the child to include animals in the sexual abuse, which can cause additional psychological harm to the child. The priority offences of child sexual exploitation and abuse (CSEA), including grooming and CSAM offences, are explored in the 'Child Sexual Exploitation and Abuse (CSEA)' chapter.

17.10 There is also some evidence that viewing acts of animal cruelty can be a precursor to the viewer performing those acts themselves, or that viewing or performing acts are an indicator for that person potentially 'graduating' to other illegal acts in the future, such as murder and child abuse.[1881] The research on these links is not specific to the posting or viewing of content *online*, nor on viewing content that encourages, assists or conspires to commit acts of animal cruelty. However, the evidence presented in this section demonstrates that this type of content is available online. As such, it cannot be discounted that this is one way in which someone may be exposed to it, after which they may go on to perform illegal acts in future.

How animal cruelty manifests online

17.11 This section is an overview which looks at how animal cruelty manifests online, and how individuals may be at risk of harm.

17.12 Ofcom's Online Experiences Tracker suggests that 78% of people in the UK are highly concerned about animal cruelty content online.[1882] The RSPCA have reported that animal cruelty cases are rising in number.[1883]

17.13 Although there is not clear evidence for a corresponding increase in *online content*, especially that which encourages, assists or conspires to commit acts of animal cruelty, we

[1881] This subject has been the subject of academic discourse for many years. There are several older studies on the subject including Thompson, K.L. and Gullone, E. 2006. An investigation into the association between the witnessing of animal abuse and adolescents' behaviour toward animals, *Society & Animals* 6, 221-243 [accessed 24 June 2024]; McVie, S. 2007. Animal abuse among young people aged 13 to 17. *Royal Society for the Prevention of Cruelty to Animals/University of Edinburgh*. [accessed 24 June 2024].
More recent research has also drawn links between a child viewing animal abuse or other types of familial violence and going on to abuse humans or animals themselves, such as Jegatheesan, B., Enders-Slegers, M-J., Ormerod, E. and Boyden, P. 2020. Understanding the link between animal cruelty and family violence: the bioecological systems model, *International Journal of Environmental Research and Public Health* 17. [accessed 24 June 2024].
[1882] Ofcom, 2024. Online Experiences Tracker. [accessed 18 November 2024]. Fieldwork was carried out in a four-week period in May and June 2024. Note that in our November 2023 and August 2024 Consultations we referred to previous iterations of this research (Waves 4 and 5).
[1883] RSPCA, 2023. Cruelty to cats increased by 25% on last year; RSPCA, 2023. One dog abused every hour: cruelty to dogs on the rise [both accessed 24 June 2024].

believe it is a reasonable assumption that as the prevalence of an offence increases generally, the online manifestation of that offence is likely to rise with it. The National Wildlife Crime Unit in the UK notes that the internet has enabled various forms of wildlife crime, including the promotion or facilitation of illegal activities such as hare coursing competitions.[1884]

Risks of harm presented by animal cruelty

17.14 Animal cruelty content takes various forms. The most extreme examples involve physical and psychological torture, bestiality and sexual abuse, live animals being crushed or eaten, fighting events, and gratuitous imagery of wounded and dead animals.

17.15 For some animal content online, users may not recognise mistreatment. 'Fake rescue' videos are manufactured scenarios in which an animal is put in a harmful situation and then someone is filmed "rescuing" it, such as a predator and prey animal being introduced and the prey animal then being "rescued". These often involve cruelty to all animals involved and can be paired with 'fake outrage'.[1885] These play on viewers' care for animals and prompt (even implicitly) comments praising the poster for their actions.

17.16 In other cases, animals may be shown being kept as pets in unsuitable conditions (especially wildlife as opposed to domesticated animals), being dressed up and forced to behave like humans (e.g. walking on hind legs, dancing, or acting like a baby), or being put in situations where they show signs of fear or distress (but with behaviour which is not widely understood as fear or distress, such as monkeys "grinning" or slow lorises "laughing"). In these scenarios, it is possible that the content creators themselves are not aware of the cruelty and do not intend to harm the animal. However, if they ought reasonably to be aware, they would be committing the animal cruelty offence.

17.17 This section will focus on evidence for the risk factors of online services that can contribute either to the priority offences explained above or to the use of the service for their facilitation or commission. However, it will likely also cover evidence that relates to more general animal cruelty content online. This may include content which would also amount to the separate non-priority 's.127(1) offence', which is covered by the Register of Risks chapter 'Obscene content showing torture of humans and animals (the s.127(1) offence)'.

17.18 We acknowledge that pre-recorded content, including where it is not openly encouraging, assisting or conspiring to commit acts of animal cruelty, or not showing obscene torture, can serve to normalise certain treatment of animals. We also recognise that this could lead to animal cruelty content that *would* constitute the priority offence, or lead to users seeking out this type of content and networks of bad actors with whom to conspire.

17.19 Therefore, while not all animal cruelty content will meet the threshold for the priority or non-priority offences, we will consider the risk factors on the basis that a broader category of animal and animal cruelty-related content online may inherently create demand for

[1884] National Wildlife Crime Unit [date unspecified]. *Cyber enabled wildlife crime*. [accessed 9 May 2024].
[1885] Harrington, L.A., Elwin, A., Paterson, S. and D'Cruze, N. 2023. 'The viewer doesn't always seem to care – response to fake animal rescues on YouTube and implications for social media self-policing policies', *People and Nature* 5 [accessed 24 June 2024]; World Animal Protection, 2021. Views that abuse: the rise of fake "animal rescue" videos on YouTube; Social Media Animal Cruelty Coalition, 2024. Spot the Scam: Unmasking Fake Animal Rescues. [accessed 28 October 2024].

more of the same, and that its existence can therefore contribute to activities which *do* encourage, assist or conspire.

17.20 Viewing illegal animal cruelty content could pose risks of harm to users online: in the case of the animal cruelty offence, viewing content which encourages, assists or conspires to animal cruelty may distress a user, or alternatively lead to them engaging in harmful or illegal behaviours themselves. There may also be a cumulative impact from users being repeatedly exposed to animal cruelty-related content, not just to content which would constitute the offence.

17.21 The information presented in this section mostly does not provide evidence for animal cruelty online as an issue specific to the UK or UK users of online services. As such, this section will consider animal cruelty online primarily as a global phenomenon.

17.22 When discussing animal cruelty content online, we primarily refer to text, images and videos (livestreamed and/or on-demand). This may be on social media sites, video-sharing services and messaging services. We acknowledge that while sites that facilitate payments for acts of animal cruelty (or for content depicting these acts) are unlikely to host content that directly constitutes the animal cruelty offence, payments services are part of the wider ecosystem contributing to the issue.

Evidence of risk factors on user-to-user services

17.23 We consider that the risk factors below are liable to increase the risks of harm relating to animal cruelty content, and to committing animal cruelty. This is also summarised in the box at the start of the section.

Risk factors: Service types

17.24 Research indicates that the following types of services are used to facilitate or commit offences related to animal cruelty.

Social media and video-sharing services

17.25 Evidence shows that animal cruelty content can often be shared on social media and video-sharing services. For instance, the Social Media Animal Cruelty Coalition (SMACC) has published research which identified channels on large social and video-sharing services – some of which have thousands or even millions of subscribers/followers – which share animal cruelty content that can get a significant numbers of views.[1886] SMACC has also found that bestiality and animal sexual abuse content exists on these types of services.[1887]

[1886] Social Media Animal Cruelty Coalition, 2021. Making money from misery, p.26 [accessed 24 June 2024]. Note that SMACC's publications referenced across this Register of Risks section, and the next section on the obscene torture content offence, are among the very few systematic studies into animal-cruelty related content online, some of which may constitute the animal cruelty or obscene content offence. We refer to these sources as evidence that potentially harmful and illegal content exists online, not as an indicator of the quantity or prevalence of this content. This caveat also applies to the research produced by other charities and non-profit organisations, such as the RSPCA, Born Free USA, Lady Freethinker and the Alliance to Counter Crime Online.

[1887] As reported directly to Ofcom by the Social Media Animal Cruelty Coalition based on recent, as-yet-unpublished data from April 2024; also Social Media Animal Cruelty Coalition 2023. The cruelty you don't see, pp.51-52. [accessed 24 June 2024].

17.26 Researchers have documented certain types of animal-related content which can be cruel – such as crushing videos, fights and 'fake rescues' – on video-sharing services, some gaining millions of views.[1888]

17.27 The UK Safer Internet Centre has also reported an increase in reports of animal abuse content across 2023 and 2024.[1889]

17.28 Recent research published jointly by the RSPCA, Scottish SPCA and the Ulster SPCA showed that 43% of 16-17 year olds and 22% of adults had witnessed animal cruelty content online (including on social media services).[1890] Ofcom's own research (including the Online Experiences Tracker) and the other evidence that we note in this section also indicates that the types of content which may constitute an offence appears still to be quite widespread online and on social media.

17.29 An investigation conducted by the BBC in 2023 revealed an online community operating across video-sharing and messaging services to encourage, assist and conspire to commit and share acts of cruelty towards animals, including members sending ideas of cruel acts that could be filmed and shared.[1891]

17.30 Evidence for videos of organised dog fighting or promotion of fights has been found on social media and video-sharing services.[1892] Similarly, imagery of dogs killing wildlife such as foxes and badgers has been identified on social media sites, including a case involving a UK user.[1893]

17.31 Social media and video-sharing service accounts which post content showing cruel cosmetic modifications of pets such as tutorials, images and pet advertisements may not just normalise these practices but actively encourage others to repeat the same procedures on their own pets.[1894] This includes declawing cats, tail docking and ear cropping or 'posting' (the processes of cutting and taping dogs' ears to give them a more pointed shape), which are all procedures deemed illegal under Section 5 of the Animal Welfare Act 2006.

[1888] Carvalho, A.F., de Morais, I.O.B., and Souza, T.B. 2023. 'Profiting from cruelty: digital content creators abuse animals worldwide to incur profit', *Biological Conservation* 287; World Animal Protection, 2021. Views that abuse: the rise of fake "animal rescue" videos on YouTube. [accessed 24 June 2024].

[1889] UK Safer Internet Centre, 2024. UK Safer Internet Centre sees concerning rise in animal abuse content. [accessed 24 June 2024].

[1890] RSPCA, 2024. Kindness Index Report 2024. [accessed 20 July 2024]. The data showing that a significant number of young people view animal cruelty content online appears to reflect older research from the RSPCA, which found that 48% of the surveyed 10-18 year-olds had witnessed animal cruelty, of whom almost a quarter (23%) had seen it on social media: RSPCA, 2018. *The RSPCA's Generation Kind*, p.2, and footnote 2 [accessed 9 May 2024]. In a related document published at a similar time, they found that one in three 10-15 year-olds reported seeing animal cruelty on social media: RSPCA, 2018. *Building a kinder generation*, p.7. [accessed 9 May 2024].

[1891] BBC, 2023 Monkey Haters (documentary) [accessed 27 March 2024].

[1892] Montrose, Kogan and Oxley. 2021. The role of social media in promoting organised dog fighting, *The Veterinary Nurse*. [accessed 24 June 2024].

[1893] Mitchell, J, 2023. Gamekeeper who filmed animal fights for TikTok spared jail, Sky News, 12 December. [accessed 28 October 2024].

[1894] Heaney, P, 2021. Dog Mutilation: Breeders Cropping Ears to Follow Social Media Trend, BBC News, 13 December. [accessed 3 October 2024].

Messaging services, discussion forums and chat room services

17.32 Messaging services, particularly those with encrypted messaging, can be used to share animal cruelty content. This is because perpetrators can use these services to communicate easily, keeping their messages private and avoiding detection by the services or authorities.

17.33 Users looking to encourage and facilitate acts of animal cruelty or conspire with other individuals may provide hyperlinks to private messaging services in public content. Members of the monkey torture ring investigated by the BBC in 2023 posted videos on a video sharing service but links in video descriptions encouraged viewers to move to private messaging services to get more extreme animal cruelty content.[1895] Lady Freethinker noted that users of online social media services involved in monkey torture networks were also facilitating this content being shared in private groups on messaging and forum services.[1896]

17.34 The Scottish SPCA, in their response to our August 2024 consultation, indicated that they see examples of bad actors using messaging services to share advice on ear cropping and other mutilations of dogs,[1897] and the RSPCA similarly reports increasing use of messaging services to provide advice around hunting, including showcasing hunting dogs and their owners.[1898]

17.35 SMACC have noted that 'hopping' across several online services is an emerging trend from networks of bad actors engaging in animal cruelty-related activities.[1899] They may do so to avoid detection or to ensure the continuity of their community should they be removed from one service. This includes using standalone messaging services or messaging surfaces on social media sites to communicate with others.

File-storage and file-sharing services

17.36 We believe that the ability to share and store animal cruelty content, perhaps with the intention of acquiring more derived from future acts of cruelty, is a likely risk factor for this type of content. File-sharing services are known to be used to share and store other types of illegal content (such as child abuse imagery and terrorist content) for personal use or distribution.[1900]

17.37 Links to file-sharing services may be shared via other online services, including social media and messaging services. However, we are not currently aware of direct evidence for file-sharing services being used to do so in the case of animal cruelty content which would constitute the offence, other than where respondents to our November 2023 Consultation flagged file-sharing services as a potential risk factor.[1901]

[1895] BBC, 2023. Monkey Haters (documentary). [accessed 27 March 2024].
[1896] Lady Freethinker, 2021. YouTube 'monkey haters' form private group where members are paying to have baby monkeys tortured and killed on camera [accessed 28 June 2024].
[1897] Scottish SPCA response to August 2024 Illegal Harms Further Consultation, p.2.
[1898] RSPCA response to August 2024 Illegal Harms Further Consultation, p.1.
[1899] As reported directly to Ofcom by the Social Media Animal Cruelty Coalition based on recent, as-yet-unpublished data from April 2024.
[1900] Our analysis suggests that file-storage and file-sharing services post a particularly high risk of disseminating CSAM, terrorist content and non-consensual intimate imagery (as part of intimate image abuse) – this is covered in the relevant chapters in our Register of Risks.
[1901] Born Free Foundation response to November 2023 Illegal Harms consultation, p.3.

User-to-user pornography services

17.38 We have noted evidence above that bestiality content has been found on social media services. A study of extreme pornography cases in England and Wales between 2015 and 2017 found that the most commonly charged category was that of extreme pornography involving an animal.[1902] As such, it is likely that this is reflected to some extent online, and specifically on online user-to-user pornography services.

17.39 The evidence and risk factors for extreme pornography – including bestiality – are explored in the 'Extreme pornography' chapter of the Register of Risks.

Risk factors: User base

User base size

17.40 Animal cruelty content, or which encourages, assists or conspires, appears on services with both large and small user bases. Different user base size can pose different risks.

17.41 The larger a service's user base, the greater the number of people who are likely to encounter content on it, meaning that content can receive substantial amounts of engagement. As shown through the evidence presented in this section (including that which was collected by SMACC), animal cruelty content exists on some of the largest user-to-user services and appears to be largely shared openly.

17.42 Bad actors may also post on larger services less-obviously cruel content or content which is not openly encouraging animal cruelty. This may desensitise viewers to seeing animals in unsuitable conditions and being treated inappropriately,[1903] thereby potentially emboldening the bad actors to post more extreme content with the same wide distribution in future.

17.43 Also, by posting content with descriptions containing external links on larger services, they can funnel viewers through to smaller and more private services or parts of services, such as messaging services and private and invite-only groups.[1904] Some of these groups may be communities *within* larger services which, while small, are still publicly accessible and have few barriers to users joining.[1905]

17.44 The smaller a provider's user base, bad actors may be incentivised to publish the most extreme content there from the start, or to organise and conspire with other like-minded people. Smaller groups of users can more easily share tips on how to avoid detection by the service's moderation systems – this has been evidenced for bad actors engaged in the illegal wildlife trade in specific-purpose groups or pages on social media services, many of which were relatively small (and in some cases, private),[1906] and it seems reasonable that this could also hold for animal cruelty content.

[1902] McGlynn, C. and Bows., H. 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, The Journal of Criminal Law, 83(6), pp.481-482. [accessed 24 June 2024].
[1903] Born Free USA, 2022. Their lives for your likes: the exploitation of wild animals on social media [accessed 10 May 2024].
[1904] BBC, 2023. Monkey Haters (documentary) [accessed 27 March 2024].
[1905] As reported directly to Ofcom by the RSPCA: Ofcom/RSPCA meeting, October 2023, and by the Social Media Animal Cruelty Coalition based on recent, as-yet-unpublished research from April 2024, which found that individuals (including children) can easily join private groups or forum by requesting access.
[1906] Alliance to Counter Crime Online, 2020. Two Clicks Away: wildlife sales on Facebook, Appendix B, which documents the Pages and Groups identified, including membership sizes [accessed 8 May 2024].

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles and fake user profiles

17.45 Anonymous user profiles can help bad actors evade detection, allowing them to encourage, assist or conspire to commit illegal acts without being identifiable. We believe it is reasonable to presume that anonymous user profiles increase the risks of users posting and engaging with animal cruelty content.

User networking

User connections

17.46 We believe that this is a risk factor for animal cruelty related harms online. The ability for users to connect with one another allows them to discuss ways in which animals may be harmed, and to share imagery showing these acts, or engage with trade relating to animals, animal products and animal cruelty content.[1907]

17.47 Users do not need to be directly connected to each other to see content posted (where it is in public feeds, for instance), but users that are connected, particularly through group messaging, are more enabled to discuss and conspire to commit acts of animal cruelty together – see the following paragraphs.

User groups and group messaging

17.48 Animal cruelty activity and content can appear in public group spaces. For instance, a Lady Freethinker study identified over 150 user groups and pages on a social media service (which could be searched within the service itself) centring around dog-fighting – including the facilitation of this activity through selling of dogs, encouraging it through content captions, and conspiring to commit animal cruelty through promoting or depicting fights.[1908]

17.49 However, some of the illegal animal cruelty content exists within closed groups and messaging services. Closed groups and group messages allow perpetrators to form a community around a shared interest and, for example, discuss ideas for acts of animal cruelty or encourage others to engage with this content.[1909] [1910]

17.50 The monkey torture network investigated by the BBC was found to have used a polling function within group messaging services to brainstorm ideas for acts of cruelty, which

[1907] Being able to accrue large followings may build a sense of legitimacy (such as on accounts sharing 'fake rescue' content and requesting financial donations), a potential risk factor for fraud. See the Fraud and financial services chapter of the Register of Risk.
[1908] Lady Freethinker 2019. The deadly, underground world of dogfighting on Facebook [accessed 8 May 2024].
[1909] BBC, 2023 Monkey Haters (documentary) [accessed 27 March 2024]; also, as reported directly to Ofcom by the RSPCA: Ofcom/RSPCA meeting, October 2023.
[1910] In one case prosecuted by the Crown Prosecution Service a user joined a private group on social media, posted videos and wrote comments that showed approval, therefore potentially encouraging engagement with other users. This case was prosecuted under the Obscene Publications Act 1959 – see our Illegal Content Judgements Guidance paragraph xxx on our decision to use s127(1) of the Communications Act 2003 rather than the Obscene Publications Act. Crown Prosecution Service, 2024. Man jailed for posting videos of baby monkeys being tortured. [accessed 25 October 2024].

were then used to encourage their connections around the world to create this type of content.[1911]

User communication

Posting content (images and videos)

17.51 The ability for users to post content, in particular images and videos, is a known risk factor for animal cruelty content, including that which encourages, assists or conspires. Posting content can allow bad actors to share animal cruelty content with a potentially large number of users. Research by various organisations has identified accounts video-sharing and social media services, and across other social channels, which publicly share animal cruelty content that can get significant numbers of views.[1912]

17.52 Content could encourage, assist or conspire to commit acts of animal cruelty, such as publicising upcoming animal fighting events (which may or may not be recorded and uploaded, or livestreamed) or clearly showing approval for the cruelty depicted. This could then perpetuate an environment in which these behaviours are encouraged and users of online services encourage the creation of more content like this.

17.53 However, even if it does not directly encourage, assist or conspire, the publishing of any animal cruelty content may normalise harmful behaviours towards animals and desensitise users of online services to these acts. This may then itself result in users encouraging increasingly extreme acts of animal cruelty to be carried out in order to be published online.

Re-posting and sharing/forwarding content

17.54 The ability to repost or forward content may be a risk factor for the animal cruelty offence in that bad actors may re-share content with others within their network, as a way to show approval for the acts and encourage further cruelty. This could, in some cases, meet the threshold for the animal cruelty priority offence. Other users may share content because they want to demonstrate their outrage or speak out *against* this type of content.

17.55 There is also the risk that users inadvertently re-post or forward animal cruelty content because they do not realise what it is. They may think the content is cute, funny, or depicts an animal being rescued, and therefore want to share it, perhaps leading to it going viral.[1913] However, bad actors may in part rely on this to achieve a wider audience for their content, normalising the behaviour, and encouraging further acts of animal cruelty.

Livestreaming

17.56 Livestreaming functionality appears likely to give rise to a risk of animal cruelty content being disseminated. Livestreaming acts of animal cruelty is likely to amount to illegal content in that it is a conspiracy between the viewers and the organisers to commit animal cruelty. Respondents to our November 2023 Consultation noted that they believed

---

[1911] Lawson, E., Gunter, J. and Henschke, R., 2024, Kidderminster women pleads guilty to role in monkey torture network, BBC News, 7 May. [accessed 8 May 2024].

[1912] Social Media Animal Cruelty Coalition, 2021. Making money from misery [accessed 24 June 2024]; South West Grid for Learning (SWGfL), 2024. 'Report Harmful Content sees concerning rise in animal abuse content' [accessed 8 May 2024]; Mitchell, J, 2023. Gamekeeper who filmed animal fights for TikTok spared jail, Sky News, 12 December. [accessed 28 October 2024]; World Animal Protection, 2021. Views that abuse: the rise of fake "animal rescue" videos on YouTube. [accessed 31 October 2024].

[1913] On fake rescue videos: Social Media Animal Cruelty Coalition, 2024. Spot the Scam: Unmasking Fake Animal Rescues. [accessed 28 October 2024]; World Animal Protection, 2021.

livestreaming to be a risk factor, which could potentially include animal fights and torture.[1914] There may be an element of livestreaming of cruelty within animal torture groups, such as the cat torture networks originating in China.[1915] Cooking or eating videos are another specific genre of content which may show the unnecessary suffering of animals, albeit with limited evidence for this content being livestreamed.[1916]

17.57 While typically ephemeral in nature, the streamer or any of the viewers may screen-cap or record the stream with the intention of posting it later on other channels and encouraging others to engage with the material – this has been shown to be a risk factor for other illegal activities, such as terrorism, and we believe it is reasonable to infer that this could also be true for animal cruelty offences.

17.58 Similarly, comment functions on streams is a known risk factor for other illegal activities, such as child grooming and CSAM, where viewers can communicate with those carrying out the illegal activities. Given the potential risk of livestreamed animal cruelty content, it is likely also a risk factor allowing users to conspire to commit cruelty, or to encourage or assist its commission by urging on, or suggesting or requesting specific acts, in real-time.

Direct messaging and encrypted messaging

17.59 We have noted above that messaging services are a risk factor for the animal cruelty offence. Two aspects of messaging functionalities could contribute to this risk – the ability to contact other online service users directly, and encryption.

17.60 Direct messaging is a risk factor for various aspects of animal abuse because people may post in public channels and through that content encourage viewers to follow-up via direct messaging or through messaging services. This builds a network of bad actors who together perpetuate acts of animal cruelty.

17.61 The additional privacy offered by encryption - applying to all interactions on a messaging service, or just to specific messaging features within a social media or messaging service – may facilitate bad actors to engage in animal cruelty activities with others, without fear of detection.[1917] For instance, the Social Media Animal Cruelty Coalition has observed that people conspiring to arrange dog fights were using encrypted groups on a social media service.[1918]

Commenting on content

17.62 Comment functions are a clear risk factor for the animal cruelty offences. As noted above for livestreamed content, users may encourage acts of animal cruelty – either in the moment or for the creation of new content. They may use this functionality to request

---

[1914] For example, the RSPCA response to November 2023 Illegal Harms Consultation, p.2 and the Scottish SPCA response to November 2023 Illegal Harms Consultation, pp.2 and 4.
[1915] Cheung, R, 2023. A man said he'd adopt cats and torture them in a livestream. Then vigilantes took action, Vice, 9 March 2023. [accessed 24 October 2024].
[1916] In one case reported in the media (Harris, M., 2020, 'A YouTuber with over 3 million follower responded to backlash', Business Insider, 17 April. [accessed 9 May 2024]), a creator of 'mukbang' content (a type of cooking and eating video that can aim to share cultural traditions, but also overlaps with "ASMR" videos) was criticised for posting videos of herself eating live animals. In this case, the videos themselves were *not* livestreamed. However, the genre of 'mukbang' is known to be either pre-recorded or livestreamed, indicating that the latter could be risk factor for the animal cruelty offence.
[1917] National Wildlife Crime Unit [date unspecified]. *Cyber enabled wildlife crime*. [accessed 9 May 2024].
[1918] Social Media Animal Cruelty Coalition, 2021. Making money from misery, p.21 [accessed 24 June 2024].

more content of the same type, thereby implicitly validating and enabling the poster to perform more acts of cruelty and/or share more content depicting these acts.

17.63 A recent study into engagement with images of animal abuse on social media services (including comments as well as likes and sharing or replicating the content) found that respondents who appeared to show traits of callousness or dismissiveness towards animals were more likely to engage positively with animal abuse content. This could indicate that comments functionalities could be used by those interested in animal cruelty content to show their approval and encourage more of this activity.[1919]

17.64 As mentioned briefly above in the section on re-posting/sharing as a risk factor, users may engage with content – including through comments – because they do not realise what it is. A characteristic of animal welfare is that it is not always obvious to an onlooker or viewer where an animal is in distress, or the act being performed is causing unnecessary suffering. This is in part because users may not understand animal behaviour, or because the act of cruelty happens off-screen (such as poor animal welfare standards[1920], fuelling wildlife trade and poaching, 'rescue videos', illegal pets[1921] and 'teasing as torture'[1922]).

17.65 Taking this into account, comments functions may be a risk factor for perpetuating and normalising animal cruelty. Where an online service allows comments on a piece of content, this may mean that some users' comments appear to be supporting a situation in which the animal is suffering. However, it is unlikely that the majority of commenting users *intend* to encourage that suffering. An analysis of comments on videos showing exotic and endangered animals showed that there is typically a positive overall sentiment toward the content, showing that users do not always understand that these animals should not be kept or treated in the way shown.[1923] While these types of comments would not meet the threshold for illegality, it shows that comments functions could be a contributory risk factor to an offence.

17.66 A Born Free USA report provided evidence of videos showing exotic animals in captivity, including one in which the handler repeatedly provokes a python who shows signs of distress. The individuals also laugh and are thus shown to be encouraging this behaviour.[1924] As of June 2024, this video is still available, with comments enabled (the Born Free USA report does not refer to or analyse these, however). Where an individual could be reasonably expected to recognise that their actions are cruel, positive comments on their

---

[1919] Conversely, the study found some correlation between participants who said they would comment to show their disapproval and higher levels of self-esteem (in that these participants felt they were furthering animal welfare causes). Note this study is in pre-print (not yet fully published) and used a non-representative sample through a survey of self-reported attitudes. We therefore use it as an indicator for the potential for commenting on comment to be a factor in encouraging, assisting or conspiring to commit animal cruelty content, rather than as definite evidence that this occurs. Source: McGuirk, L.Ryan and Alleyne, E. 2024 '"Liking," "Commenting," and "Reposting": Psychological factors associated to online animal abuse', *Society & Animals*, p.17. [accessed 18 November 2024].

[1920] Social Media Animal Cruelty Coalition 2023. The cruelty you don't see. [accessed 24 June 2024].

[1921] Social Media Animal Cruelty Coalition, 2022a. Wild animal "pets" on social media. [accessed 24 June 2024].

[1922] Social Media Animal Cruelty Coalition, 2022b. Teasing as Torture. [accessed 24 June 2024]. The positive (if apparently short-lived) impact of the 2015 'Tickling as Torture' campaign which aimed to raise public awareness of an act which is not clearly animal cruelty is referenced in Moloney, G.K., Tuke, J., Dal Grande, E., Nielsen, T. and Chaber, A.-L. 2021. 'Is YouTube promoting the exotic pet trade? Analysis of the global public perception of popular YouTube videos featuring threatened exotic animals'. [accessed 24 June 2024].

[1923] Moloney, G.K. et al. 2021. [accessed 24 June 2024].

[1924] Born Free USA, 2022, p.11. [accessed 24 June 2024].

content may nevertheless motivate them to continue performing these acts. A service with comments functions therefore could facilitate the animal cruelty offence.

17.67 SMACC has also observed that bad actors can use coded or ambiguous language to evade detection.[1925] This is potentially more an issue within public channels, such as comments sections. They may also be able to share hyperlinks to other material or to other services on which more extreme material is shared.

Transaction and offers

Posting goods and services for sale

17.68 The monkey torture ring uncovered by the BBC also involved members of the network were paying for people in other countries to create the content.[1926] The torture 'services' were not necessarily 'for sale', but each requested video was, in effect, an advert that encouraged further payments for future acts.

17.69 While we are not currently aware of evidence to suggest payments or rewards systems on livestreaming services is a particular risk factor, in the case of animal cruelty it is reasonable to assume that this could directly motivate further acts of animal cruelty in real-time, therefore directly facilitating the animal cruelty offence.

17.70 Where an online service provides integration to an online shop or marketplaces, or even payment services on an off-platform, this may enable users who are interested in animal cruelty content to conduct transactions easily, while also incentivising the suppliers to encourage sales of existing or future content.[1927] Payment systems, or the ability to share links to external payments system, could therefore contribute to the commission and facilitation of animal cruelty.

Recommender systems

Content recommender systems

17.71 Bad actors may take advantage of content recommender systems that prioritise user engagement by posting animal cruelty content and hoping it will get wide reach. This may constitute the priority offence where there is intent to encourage, assist, or conspire to further acts of animal cruelty. In particular, even those not actively seeking out or wishing to engage with animal cruelty and wildlife trade content may be recommended it because they have previously watched or sought out other animal content.[1928] If a user then views

---

[1925] Social Media Animal Cruelty Coalition, 2021. Making money from misery, p.21. [accessed 24 June 2024].

[1926] Two UK women charged with encouraging or assisting unnecessary suffering made payments for content via PayPal, as reported in the media: Lawson, E., Gunter, J. and Henschke, R., 2024; and Rack, S. and Cooper, M, Monkey torture video accused woman granted bail, BBC News, 11 June [accessed 8 May 2024]. In the broader international network, individuals charged in the US collected money from group members and sent it to their contact abroad who then created the videos: Office of Public Affairs (US Department of Justice), 2024. 'Two charged for involvement with online groups dedicated to monkey torture and mutilation'. [accessed 22 October 2024]; US Department of Homeland Security, 2024. 'Virginia man pleads guilty to producing, distributing sadistic animal torture videos following HSI Norfolk investigation'. [accessed 22 October 2024].

[1927] Social Media Animal Cruelty Coalition, 2021. Making money from misery, p.45. [accessed 24 June 2024]. From the context of the wildlife trade, there is some evidence of pages providing options for admins to use shopping feature and allow users to easily send direct messages to sellers (Alliance to Counter Crime Online, 2020. Two Clicks Away: wildlife sales on Facebook. [accessed 8 May 2024]). While this does not necessarily constitute the animal cruelty offence, it indicates it could be a risk factor for this content.

[1928] BBC, 2023 Monkey Haters (documentary). [accessed 27 March 2024].

cruelty content – whether or not they realise what it is – they may be shown more similar content.[1929]

17.72 Content recommender systems can also amplify animal cruelty content even where users react negatively to or comment on it to indicate their concern, anger or disgust at the content, for instance, as the services' automated systems may still consider this to be engagement.[1930]

Network recommender systems

17.73 Recommendations of other users may be a risk factor for connecting users who may form an online network which shares illegal animal cruelty content. There is some evidence that – within the context of the wildlife trade, at least – recommendations to pages or groups surface other bad content to those who have engaged with it in other places: during a study to identify the prevalence of wildlife sales, the Alliance to Counter Crime Online found 29% of the pages in their sample through the recommendations.[1931] We believe it is reasonable to presume that this risk factor could also apply in the case of connecting users engaged in animal cruelty activities.

Content exploring factors

User generated content searching

17.74 Given that the existence of animal cruelty content online in itself could encourage or represent a conspiracy to commit animal cruelty, the ability of users to *search* for this content within a user-to-user service, or for it to appear in search results (even if not directly searched for) is a risk factor for the animal cruelty offence.

17.75 Studies conducted by research bodies and non-profit organisations have shown that animal cruelty content and communities which share it (such as groups or pages), can be easily found on social media and video-sharing services through the search functions within those services. The method used for these studies included collecting data found from key word searches on video-sharing, social media and messaging services.[1932]

Risk factors: Business models and commercial profiles

Business model (revenue model and growth strategy)

17.76 Online provider's business models often give rise to financial incentives to maximise user engagement. In some cases, online services can be incentivised to recommend content that is illegal or harmful, or to facilitate sharing of such content, if it is engaging for certain communities of users.

[1929] The Social Media Animal Cruelty Coalition has told Ofcom that when conducting a specific piece of research on "fake rescue" content, 20% of the 763 pieces of content that they recorded had been recommended to their volunteer researchers' personal accounts as opposed to having been searched for; while it is difficult to assess whether this is widespread amongst other users, it suggests that this could be a risk.

[1930] The Social Media Animal Cruelty Coalition advises users not to engage with animal cruelty content, noting that views and engagement may increase popularity of the view. For example: Social Media Animal Cruelty Coalition, 2021, p.14. [accessed 24 June 2024].

[1931] Alliance to Counter Crime Online, 2020.

[1932] Carvalho, A.F., de Morais, I.O.B., and Souza, T.B. 2023; Social Media Animal Cruelty Coalition, 2021. Making money from misery, p.16. [accessed 24 June 2024]; Social Media Animal Cruelty Coalition 2023. The cruelty you don't see, p.12. [accessed 24 June 2024].

17.77 There is some evidence that animal cruelty content could drive engagement with some users, as the videos identified by researchers can get thousands or millions of views. A study of over 400 videos which featured activities which may show unnecessary suffering of animals (such as hunting experiments, slaughtering, 'fake rescues' and fights) showed that 28% were monetised. While not every category of video was monetised (the researchers did not identify any monetised 'crushing' videos from their sample), nor monetised equally, this nevertheless suggests that a variety of animal-related videos – including where they may constitute the animal cruelty offence, can make money for both content creator and online service.[1933]

17.78 The same study also noted that the content creators producing 'fake rescue' videos can ask for donations, indicating that content creators may have incentives to continue misrepresenting their activities to make money.[1934]

Commercial profile

17.79 We are not currently aware of any evidence to suggest that the size or the stage of development of a service is a particular risk factor for the animal cruelty offence.

[1933] Carvalho, A.F., de Morais, I.O.B., and Souza, T.B. 2023.
[1934] Carvalho, A.F., de Morais, I.O.B., and Souza, T.B. 2023, pp.3 and 8; see also Social Media Animal Cruelty Coalition, 2024. Spot the Scam: Unmasking Fake Animal Rescues, pp.17 and 28-29. [accessed 28 October 2024].

18. Epilepsy trolling offence

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for epilepsy trolling offence: how harm manifests online and risk factors

Some individuals with epilepsy may have a physical reaction to online content; they may feel disorientated, uncomfortable or unwell after seeing certain images or patterns. The offence covered in this chapter is sharing an image with the intention to cause harm to an individual with epilepsy. As the epilepsy trolling offence is new, it is difficult to state the prevalence of the harm. However, Epilepsy Action, a British charity, states that around one in 100 people in the UK have epilepsy. Of these individuals, 3% have photosensitive epilepsy.[1935]

The risks of harm to individuals – both adults and children – from epilepsy trolling can be psychological and physical. Some individuals may lose their life after having a seizure. Targeted attacks can also cause anxiety among epileptic users.

Service type risk factors:

Social media services have been identified as having higher risks of harm connected to epilepsy trolling, although any service which allows upload of images or videos may be a risk.

User base risk factors:

Disability and age are risk factors for this offence. Charities supporting **people with epilepsy** advise that they are being targeted, and that **young people** may be particularly vulnerable.

Functionalities and recommender systems risk factors:

Evidence indicated that perpetrators can create multiple **user profiles** to evade account-blocking efforts, and often carry out epilepsy trolling anonymously, such as by creating **anonymous user profiles**. The **user connections** displayed on user profiles can be used to find potential victims.

The ability to **comment** and **post content** such as videos or images allows perpetrators to share flashing or contrasting visual media, deliberately targeting victims and survivors. Perpetrators can **tag users** in posted content, which contains flashing or contrasting visual media, to trigger seizures. **Tagging content** with popular hashtags, can also increase the risk of content with flashing or contrasting visual media (which has the potential to trigger seizures) being disseminated on a service. **Recommender systems** can increase the risk that this content, if tagged

[1935] Epilepsy Action, n.d. Photosensitive epilepsy. [accessed 10 May 2023].

with epilepsy-related hashtags, will be seen by those engaging with epilepsy-related content (perhaps because they have epilepsy themselves).

Introduction

18.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the epilepsy trolling offence listed under 'Relevant offences' below; and
- the use of these services for the commission and/or facilitation of this offence (collectively the 'risks of harm'). [1936]

18.2 We set out the characteristics of U2U services that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.[1937]

18.3 The Epilepsy Society played a key role in its inclusion of this offence in the Online Safety Act (the Act). Due to this, a lot of the evidence included within this chapter relies on the evidence the Epilepsy Society submitted to the Government, as well as the engagement we had with them to understand the risks of harm of epilepsy trolling.

Relevant offence

18.4 In this chapter we consider the offence of sending or showing flashing images electronically, set out in section 183 of the Act. Commonly referred to as 'epilepsy trolling', this offence involves sending flashing images[1938] electronically to trigger seizures, or cause alarm or distress, among people with epilepsy, in particular those with photosensitive epilepsy.

18.5 For more details on this offence and how services can assess whether content amounts to illegal content, please refer to the Illegal Content Judgements Guidance or ICJG.

How epilepsy trolling manifests online

18.6 This section is an overview which looks at how the epilepsy trolling offence may manifest online, and how individuals may be at risk of harm. As the epilepsy trolling offence is new, we necessarily draw from evidence about behaviour that appears broadly similar to that which is intended to be captured by this new offence, recognising that the specific examples given, by definition, pre-date the new offence itself.

[1936] We have considered in other chapters how U2U services can be used to commit or facilitate priority offences, as required by Ofcom's risk assessment duty. Epilepsy trolling is not a priority offence; however, our analysis has also considered how U2U services might be used to commit or facilitate the offence, and we set out evidence how this may happen in this chapter.

[1937] We note that the epilepsy offence itself adopts a different definition of harm (see section 183(13)).

[1938] Flashing images include GIFs.

18.7 Epilepsy Action, which provides information, advice and support for people with epilepsy, states that around one in 100 people in the UK have epilepsy. Of these individuals, 3% have photosensitive epilepsy.[1939]

18.8 The Epilepsy Society notes that this type of harm is likely to take place on social media services. It states that it appears to be *"a new phenomenon born out of global communities where people can operate behind hidden identities, using new technology to provoke seizures and cause bodily harm."* [1940]

18.9 In May 2020 there was a sustained *"attack by internet trolls"*[1941] on the Epilepsy Society's X (formerly Twitter) account page, as well as through the accounts of many of its followers. It was not the first attack that had occurred; however, the Epilepsy Society stated that it did appear to be the most sustained and 'vicious'.[1942]

18.10 To provide context on how many users may have been exposed to this behaviour, the Epilepsy Society X page currently has more than 30,000 followers.[1943] It is unclear how many of these may have had photosensitive epilepsy themselves.

18.11 The Epilepsy Society told Ofcom that although followers of its social media accounts appear to have been targeted, it believes that followers of two other epilepsy charities (Young Epilepsy and Epilepsy Action) have been subjected to similar attacks.[1944]

Risks of harm to individuals presented by the offence of epilepsy trolling online

18.12 In the May 2020 series of attacks involving the Epilepsy Society's X page, perpetrators posted hundreds of flashing images and GIFs in comments on the Epilepsy Society's posts, as well as through direct messaging to some of the followers of the account, with the aim of triggering seizures.

18.13 The Epilepsy Society told Ofcom that perpetrators are sometimes strategic when selecting victims to target, intentionally timing attacks to cause seizures that would interfere with opportunities the victim may have been pursuing (and which require them to have been seizure-free for a particular period).[1945]

18.14 As well as the psychological impact that this harm can have, some individuals with epilepsy (including those with photosensitive epilepsy) may have a physical reaction to this type of content and feel disorientated, uncomfortable or unwell after seeing the images or patterns online.[1946] Seizures can also be fatal due to multiple outcomes that may result from them.[1947] This offence may also cause emotional harm by causing alarm or distress among those with epilepsy.

[1939] Epilepsy Action, n.d. Photosensitive epilepsy. [accessed 10 May 2023].
[1940] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].
[1941] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 15 June 2023].
[1942] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].
[1943] Source: Epilepsy Society, n.d. Epilepsy Society X Page. [accessed 31 July 2023].
[1944] Meeting between Ofcom and Epilepsy Society, April 2023.
[1945] Meeting between Ofcom and Epilepsy Society, April 2023.
[1946] Epilepsy Society, n.d. Photosensitive epilepsy. [accessed 2 May 2023].
[1947] Meeting between Ofcom and Epilepsy Society, April 2023.

18.15 The Epilepsy Society also told Ofcom that around the time of the attacks, those within the epilepsy community more generally felt a heightened state of anxiety when online, due to the perpetrators' intent to deliberately cause harm to individuals within the epilepsy community.[1948]

18.16 Written evidence from the Epilepsy Society submitted to Government described two cases of users who had been harmed, demonstrating that both adults[1949] and children[1950] can be affected by this harm.

Evidence of risk factors on user-to-user services

18.17 We consider that the risk factors below are liable to increase the risks of harm relating to epilepsy trolling. This is also summarised in the grey box at the start of the chapter.

Risk factors: Service types

Social media services

18.18 To date, attacks have taken place on social media services.[1951] We therefore consider social media services to be a risk factor for epilepsy trolling. However, it is possible that any service which allows users to share images, videos, GIFs, or hyperlinks to other content could also be used for epilepsy trolling.

Risk factors: User base

User base demographics

18.19 The following section outlines key evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

18.20 Evidence from the Epilepsy Society suggests that user base characteristics including age and disability could lead to increased risks of harm to an individual.

18.21 Age can be a risk factor, particularly for younger adults. The Epilepsy Society told Ofcom that photosensitive epilepsy is less commonly diagnosed over the age of 20. It noted that

[1948] Meeting between Ofcom and Epilepsy Society, April 2023.

[1949] One example was that of a 25-year-old man who had recently been diagnosed with epilepsy. He visited the Epilepsy Society X page after being recommended by friends for peer support and trustworthy information. Unfortunately, he soon came across a flashing image posted on the page which resulted in him experiencing a serious convulsive seizure which caused him to bite through his tongue. He was psychologically traumatised as a result. Source: Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].

[1950] In May 2020 a young boy aged 8 was harmed when his mum proudly shared a video of him on X of him attempting to raise money for the Epilepsy Society. Zach had epilepsy and cerebral palsy. On posts such as this, perpetrators share multiple flashing images and GIFs to try to trigger seizures in people with photosensitive epilepsy. Zach was a victim of this attack and the campaign against this behaviour was named 'Zach's Law' as a result. Source: Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].

[1951] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].

sometimes individuals with photosensitive epilepsy will experience fewer symptoms once they enter their mid to late twenties.[1952]

18.22 Disability is the main risk factor for this offence; individuals with epilepsy are those being targeted (specifically those with photosensitive epilepsy) and are at risk of suffering from negative physical effects from this content (including seizures). Epilepsy Action states that about one in 100 people in the UK have epilepsy. Of these individuals, 3% have photosensitive epilepsy.[1953]

Risk factors: Functionalities and recommender systems

User identification

User profiles and anonymous user profiles

18.23 Evidence has shown that perpetrators often create a different user profile, or multiple user profiles, once their account has been blocked by services as a result of epilepsy trolling. This can enable them to continue with the offence.[1954]

18.24 The ability to create anonymous user profiles also appears to increase the likelihood of this offence taking place. The Epilepsy Society told Ofcom that the user profiles which appeared to be engaging with this behaviour were anonymous, with false profile names and cartoon images as profile pictures.[1955]

18.25 Anonymity has been cited as one of the principal factors creating the 'disinhibition effect', allowing people to do or say things online that they would be unlikely to do if they could be identified.[1956]

User networking

User groups

18.26 User groups may be targeted by perpetrators looking for their next victims. As described by the Epilepsy Society, perpetrators targeted groups or accounts which were likely to have users with epilepsy to conduct the attacks. User groups with many followers can become a risk factor due to this.

User connections

18.27 The visibility of a user's connections may be a risk factor for epilepsy trolling. Perpetrators may use connections displayed in epilepsy-related user profiles to find other potential victims.[1957]

User tagging

18.28 The ability to tag other users in posts and comments is a risk factor for this offence. The Epilepsy Society told Ofcom that they had seen examples of perpetrators tagging the

[1952] Meeting between Ofcom and Epilepsy Society, April 2023.
[1953] Epilepsy Action, n.d. Photosensitive epilepsy. [accessed 10 May 2023].
[1954] Meeting between Ofcom and Epilepsy Society, April 2023.
[1955] Meeting between Ofcom and Epilepsy Society, April 2023.
[1956] Suler, J., 2004. The Online Disinhibition Effect, *Cyberpsychology & behaviour: the impact of the internet, multimedia and virtual reality on behaviour and society,* 7 (3). [accessed 27 April 2023].
[1957] Meeting between Ofcom and Epilepsy Society, April 2023.

Epilepsy Society's X page, as well as tagging users who had recently tweeted about their epilepsy, in the comments of posts with harmful content that could trigger seizures.[1958]

User communications

Direct messaging

18.29 As well as posting this content in public spaces on social media services, perpetrators have been known to message flashing images/GIFs to users with epilepsy. The Epilepsy Society told Ofcom that victims had received direct messages, containing content that could trigger seizures, on services such as Instagram.[1959]

Posting content and commenting on content (images and videos)

18.30 The ability to post content, in particular images and videos, increases the risk of epilepsy trolling. In screenshots provided by the Epilepsy Society of posts on a social media service, there is evidence of perpetrators posting images and videos that contain flashing or contrasting elements to users who have posted about their epilepsy.[1960]

18.31 Commenting functionalities on U2U services are a risk factor for epilepsy trolling. There have been examples of perpetrators commenting a flashing image on a user's posts. Sometimes this comment will be an image alone, but on other occasions images are accompanied by harmful text, which clearly communicates the malicious intent behind the comment.[1961]

18.32 The ability to comment on threads and posts has been used by perpetrators to exchange tips and share harmful content that may trigger seizures. In screenshots provided by the Epilepsy Society, one user had created a thread of harmful GIFs to send to users with epilepsy. Within the comments, GIFs had been shared by the user. Screenshots also showed a link to this thread being shared in the comments of a post of an individual who was celebrating a period without any seizures. [1962]

Content exploring

Content tagging

18.33 The ability to tag content through hashtags, for instance, can increase the risk of content with the potential to trigger seizures being disseminated on a platform. Hashtags enable perpetrators to share or comment on content with flashing/contrasting elements alongside hashtags associated with epilepsy. As a result, people with epilepsy who may be searching using epilepsy-related hashtags will be more likely to encounter this type of content. It also enables perpetrators to find potential victims by searching for epilepsy hashtags to find users who may have previously posted about their epilepsy.[1963]

18.34 The Epilepsy Society told Ofcom that perpetrators appear to be able to find potential victims by searching for user-generated posts with epilepsy-related hashtags.[1964]

[1958] Meeting between Ofcom and Epilepsy Society, April 2023.
[1959] Meeting between Ofcom and Epilepsy Society, April 2023.
[1960] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].
[1961] Meeting between Ofcom and Epilepsy Society, April 2023.
[1962] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].
[1963] Meeting between Ofcom and Epilepsy Society, April 2023.
[1964] Meeting between Ofcom and Epilepsy Society, April 2023.

Hyperlinking

18.35 The ability to hyperlink external content to users on a service can increase the risk of exposure to this content. Perpetrators may share links to flashing images that exist elsewhere on the internet to users with epilepsy, without them knowing the content they are being led to.

Content editing

Editing visual media

18.36 Functionalities that allow users to edit clips of visual media could increase the risk of perpetrators being able to create content to facilitate this offence. The Epilepsy Society told Ofcom that they had seen evidence of celebrity music videos being clipped, and often turned into a GIF. These would typically be of music videos that featured clips with flashing images, which would then be posted to users with epilepsy.[1965]

18.37 Editing functionalities are sometimes used by perpetrators to create malicious content. Perpetrators occasionally upload self-created content, which may include a flashing image with malicious wording (such as wording inciting their intent to cause a seizure) overlaid on the image. [1966]

Recommender systems

Content recommender systems

18.38 Evidence suggests that content recommender systems can increase the risk of unintended exposure to this type of content. The way in which recommender systems are designed can influence the extent to which illegal content is recommended to users.

18.39 Explicit user feedback on content can play an important role in the extent to which that content is disseminated across a service. Where users express interest in content through positive user feedback (such as likes, shares, and comments), recommender systems are likely to amplify that content. For example, when users with epilepsy explicitly engage in epilepsy-related content (e.g., through likes, comments, and reshares), recommender systems will learn that the user is interested in this type of content. If a user is primarily engaging with epilepsy content and not with other types of content, this is likely to create a 'filter bubble'; the user is recommended more epilepsy content while other content is deprioritised. And when perpetrators use epilepsy-related hashtags when creating or sharing this type of harmful content, victims can inadvertently be recommended this content alongside other epilepsy-related content. The Epilepsy Society told Ofcom that it was aware of users saying that they had seen this type of harmful content when scrolling through their social media feeds.[1967]

18.40 It is possible that when perpetrators tag users in content and/or use hashtags which people with epilepsy are likely to use, content recommender systems will pick this content up and distribute it more widely.[1968]

[1965] Meeting between Ofcom and Epilepsy Society, April 2023.
[1966] Meeting between Ofcom and Epilepsy Society, April 2023.
[1967] Meeting between Ofcom and Epilepsy Society, April 2023.
[1968] Meeting between Ofcom and Epilepsy Society, April 2023.

18.41 The Epilepsy Society noted that auto-play features on videos can also pose a threat, particularly where a user is recommended a video containing flashing/contrasting elements which is then played to them automatically.[1969]

Risk factors: Business models and commercial profiles

18.42 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[1969] Meeting between Ofcom and Epilepsy Society, April 2023.

19. Cyberflashing

Warning: this chapter contains content that may be upsetting or distressing.

Summary analysis for cyberflashing offence: how harm manifests online, and risk factors

The cyberflashing offence refers to the sending of a photograph or film of genitals, to cause alarm, distress or humiliation, or to obtain sexual gratification. A 2018 study showed that over 40% of women in the UK aged between 18 and 26 had been cyberflashed at some point in their lives, with rates being higher among young people. Some individuals report very high levels of cyberflashing, receiving several images a day.

The risks of harm to individuals from cyberflashing include psychological impacts. Victims and survivors describe feeling vulnerable and embarrassed, and view cyberflashing as aggressive and intimidating. In addition, cyberflashing can form part of a pattern of harmful behaviour that includes other harms such as cyberstalking, harassment, and/or controlling or coercive behaviour.

Service type risk factors:

Cyberflashing offences can occur on any service that enables users to share images. **Social media services** and **online dating services** were found to be particularly risky.

User base risk factors:

Cyberflashing is a **gendered** offence; among individuals aged 18-34, women are much more likely than men to have received an unsolicited sexual photo (40% vs 26%).[1970] There is also evidence to suggest that women in **minority ethnic groups** and **LGBT+ groups** disproportionately experience cyberflashing**.**

Functionalities and recommender system risk factors:

User connections allow perpetrators to make contact with victims and survivors. **Direct messaging,** including **ephemeral messaging**, can allow perpetrators to cyberflash victims and survivors by sending messages which contain sexual images. **Livestreaming** can facilitate a form of cyberflashing where a perpetrator exposes themselves live through a video-call.

Introduction

19.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

[1970] Sample consisted of 1629 adults. Source: YouGov (Smith, M), 2018. Four in ten young women have been sent unsolicited sexual images. [accessed 28 July 2023].

- content on U2U services that may amount to the cyberflashing offence listed under 'Relevant offences' below; and
- the use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

19.2 We set out the characteristics of U2U services that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

19.3 Cyberflashing a child under 18 years old can give rise to multiple offences and is considered a sexual communication with a child offence. Although we briefly discuss evidence that includes cyberflashing a child in this chapter, we recognise that this act in particular will overlap with a number of other offences explored in other chapters including chapter, Child sexual exploitation and abuse offences.[1971]

19.4 There are two primary mechanisms used by perpetrators to cyberflash: Bluetooth channels[1972] and online channels. Much of the existing research and evidence on cyberflashing and its prevalence does not distinguish between Bluetooth and online channels. We recognise that some U2U services may allow users to send images via both online and Bluetooth channels, and that this could enable cyberflashing. The conclusions of risk of harms to individuals in this chapter apply to all images transmitted on U2U services, regardless of the specific technology used.

19.5 There is some research and evidence available on cyberflashing via online means and its impacts, although much of it is focused on unsolicited sexual images rather than cyberflashing as it is defined in the legislation. The overlapping negative impact and prevalence however indicates an area of concern for risk of harms.

Relevant offences

19.6 In this chapter we consider the offence of cyberflashing, a new offence[1973] created by the Act.[1974] The offence of cyberflashing is committed where a person intentionally sends a photograph or film of genitals for the purposes of causing alarm, distress or humiliation or for the purpose of obtaining sexual gratification.

19.7 The Act also covers the offences of encouraging and assisting, and conspiracy to commit, this offence.

[1971] One study found that a majority of girls in a qualitative research project of 144 young people aged 11-18 had received unsolicited naked images of boys or men. Source: Ringrose, J., Regehr, K. and Whitehead, S, 2021. Teen Girl' Experiences Negotiating the Ubiquitous Dick Pic: Sexual Double Standards and the Normalization of Image Based Sexual Harassment, *Sex Roles, 85* (558). [accessed 11 August 2023].

[1972] Bluetooth allows for wireless 'pairing' between two proximate devices using a peer-to-peer network. Bluetooth 'pairing' can be used to share files between devices, and perpetrators can use this to share unsolicited explicit images with nearby devices and cyberflash the device's user.

[1973] New section 66A of the Sexual Offences Act 2003.

[1974] Section 187 of the Online Safety Act 2023.

19.8 For more details on the offences and how services can assess whether content amounts to illegal content, please refer to the Illegal Content Judgements Guidance.

19.9 Some of the evidence we have considered relates to content which may not necessarily mirror, or is broader than, the criminal definitions of this offence. This chapter therefore considers more broadly the harms from the unsolicited sending of sexual images. We refer to the sending of sexual images in this way (including images of genitals) as cyberflashing within this chapter. Like intimate image abuse, cyberflashing is a form of image-based sexual abuse and is disproportionately perpetrated by men towards women.[1975] There is some evidence that a notable proportion of men sending unsolicited genital images do so knowing it can cause harm, as they intend to provoke shock, fear and disgust in recipients.[1976]

19.10 In some circumstances, the reception of these images may be more positive. In particular, in some contexts and communities, such as specific dating platforms, the sending of unsolicited genital images may be more normalised due to different relationship dynamics or expectations.[1977]

19.11 However, when sending such images is normalised, it can also increase the risk that recipients feel pressured to accept or reciprocate genital images. Even in cases where such images are not viewed negatively, when the sender does not ask for consent before sending genital images, this fails to respect the sexual autonomy[1978] of the receiver.

How cyberflashing manifests online

19.12 This section is an overview which looks at how cyberflashing manifests online, and how individuals may be at risk of harms.

19.13 One study conducted in 2018 found that over 40% of women in the UK aged between 18 and 36 had been cyberflashed at some point in their lives.[1979] Among young people, rates of cyberflashing are even higher[1980]; 47% of women aged 18-24 have been cyberflashed[1981]

[1975] For example, in a 2019 survey of Canadian men, a notable proportion of those how confirmed they had sent unsolicited sexual images indicated "that they hoped to provoke negative emotions in recipients, with 17% hoping for shock, 15% hoping for fear, and 11% hoping for disgust." Source: Oswald, F., Lopes, A., Skoda, K., Hesse, C. and Pedersen, C., 2019. I'll Show You Mine so You'll Show Me Yours: Motivations and Personality Variables in Photographic Exhibitionism. *The Journal of Sex Research*, 57 (5), p.1-13. [accessed 07 October 2024] .

[1976] YouGov (2018), "Four in ten female millennials have been sent an unsolicited penis photo"

[1977] For example, gender-based power dynamics can be different in same sex relationships compared to heterosexual ones, resulting in different experience of sharing sexual images. Dietzel, C. 2021. The three dimensions of unsolicited Dick Pics: Men who have sex with men's experiences of sending and receiving unsolicited Dick Pics on dating apps, Sexuality & Culture, 26(3), pp. 834–852. [accessed 18 November 2024].

[1978] Sexual autonomy refers to one's right to make informed decisions related to their body, sexuality or sexual experiences, free of societal pressure and desires. Willie, T.C.., Callands, T., Alexander, K.A. 2023. Measuring women's sexual autonomy: Development and preliminary validation of the Women's Sexual Autonomy Scale, Women's Health, 19. [accessed 18 November 2024].

[1979] Sample consisted of 2,353 women aged 18-36 and 1,327 men aged 18-36. YouGov (Smith, M.), 2018. Four in ten female millennials have been sent an unsolicited penis photo. [accessed 26 July 2023].

[1980] Sample consisted of 2,353 women aged 18-36 and 1,327 men aged 18-36. 71% of women 18-24 were younger than 18 the first time they were cyberflashed. Source: YouGov (Smith, M.), 2018. Four in ten female millennials have been sent an unsolicited penis photo. [accessed 26 July 2023].

[1981] YouGov (Smith, M.), 2018.

and 19% of girls aged 11-16 have been sent unwanted sexual images.[1982] Some individuals report very high levels of cyberflashing, receiving several images a day.[1983]

Risks of harm to individuals presented by cyberflashing

19.14 Online cyberflashing can happen through any U2U services, including social media services, online dating services and video-sharing services. Perpetrators of online cyberflashing can use channels such as direct messaging and commenting to share unsolicited explicit images.[1984]

19.15 Cyberflashing is a gendered offence; the majority of victims and survivors are women and the majority of the perpetrators are men.[1985] There is also evidence to suggest that women in minority ethnic groups and LGBT+ groups disproportionately experience cyberflashing .[1986] This is part of a broader dynamic in which women, particularly minoritised women, are sexualised and fetishised in ways which harm their sexual autonomy.[1987]

19.16 The impact of harm from cyberflashing is varied and can affect individuals differently depending on the circumstances of the offence. However, there is substantial evidence indicating that cyberflashing can have psychological impacts.

19.17 Testimonies from victims and survivors of cyberflashing describe feelings of shame and embarrassment.[1988] Of women who have been cyberflashed, 58% described the unsolicited images as 'gross' and 54% described them as 'stupid'.[1989]

19.18 Victims and survivors also describe feeling vulnerable following experiences of cyberflashing.[1990] Many victims and survivors describe the experience of being cyberflashed as aggressive and intimating,[1991] and some victims and survivors describe the experience as a violation.[1992]

19.19 There is evidence that cyberflashing occurs both in cases where perpetrators are known to victims and survivors, and where perpetrators are not known to them. The harms experienced may vary depending on whether the perpetrator was previously known or not.[1993]

[1982] Sample consisted of 2,114 girls and young women aged 7-21. Source: Girlguiding, 2021. Girls' Attitudes Survey. [accessed 26 July 2023].
[1983] Revealing Reality, 2023. Anti-social Media: The violent, sexual and illegal content children are viewing on one of their most popular apps. [accessed 26 July 2023].
[1984] As mentioned above, perpetrators can also use a number of other means out of scope of the Act such as by email. We will not explore the details of these in this chapter.
[1985] Law Commission, 2021. Modernising Communications Offences: A final report. [accessed 26 July 2023].
[1986] McGlynn, C., 2021. Written evidence submitted by Professor Clare McGlynn, Durham Law School, Durham University (VAW0007). [accessed 29 July 2023]; McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. Bristol: Bristol University Press. [accessed 21 September 2023].
[1987] Glitch, UK, 2023. The Digital Misogynoir Report: Ending the dehumanising of Black women on social media. [accessed 18 November 2024].
[1988] Law Commission, 2021.
[1989] Sample consisted of 1,629 adults. Source: YouGov (Smith, M.), 2018. Four in ten young women have been sent unsolicited sexual images. [accessed 26 July 2023].
[1990] Law Commission, 2021.
[1991] The National Police Chiefs' Council response to 2020 Law Commission Consultation. [accessed 2 August 2023].
[1992] McGlynn, C. response to 2020 Law Commission Consultation. [accessed 2 August 2023].
[1993] McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. Bristol: Bristol University Press. [accessed 21 September 2023].

19.20 Cyberflashing is not a product of technology and online behaviour alone; it is a manifestation of existing patterns of sexual violence and abuse. McGlynn argues that cyberflashing should be understood as part of a continuum of sexual violence.[1994] As with all forms of sexual violence, perpetrators of this abuse are motived by a desire to exert power,[1995] and victims and survivors experience feelings of fright and vulnerability.[1996]

19.21 The cyberflashing offence can occur in conjunction with several other online harms explored in this Register. Cyberflashing can form part of a pattern of harmful behaviour that includes other harms such as cyberstalking,[1997] harassment, and controlling or coercive behaviour.[1998]

Evidence of risk factors on user-to-user services

19.22 We consider that the risk factors below are liable to increase the risks of harm relating to cyberflashing. This is also summarised in the grey box at the start of the chapter.

Risk factors: Service type

19.23 Cyberflashing offences can occur on any service that enables users to share images. Evidence suggests that social media and dating services are particularly risky service types for cyberflashing offences.[1999]

Social media services

19.24 There is evidence to suggest that cyberflashing is prevalent on social media services. Ofcom research shows that among internet users who had experienced cyberflashing in the past four weeks, 40% were using social media services at the time.[2000]

Dating services

19.25 There is also evidence that cyberflashing is particularly prevalent on dating services. A survey from dating site Plenty of Fish found that 48% of single adults who use dating sites had previously received unsolicited nude images on these sites.[2001] Often victims and survivors describe receiving images seemingly at random, but there is also evidence that cyberflashing is used as a response to the romantic rejection of the perpetrator, sometimes accompanied by other threats.[2002]

[1994] McGlynn, C., 2022. Cyberflashing: Consent, Reform and the Criminal Law, *The Journal of Criminal Law,* 85 (5). [accessed 28 July 2023].

[1995] McGlynn, C., 2021. Written evidence submitted by Professor Clare McGlynn, Durham Law School, Durham University (VAW0007). [accessed 28 July 2023].

[1996] Law Commission, 2021.

[1997] McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. Bristol: Bristol University Press. [accessed 21 September 2023]; Brazier, T., 2022. Emily Atack to front new BBC documentary about online sexual harassment after being targeted 'from a very young age', *Metro,* 4 September. [accessed 12 September 2023].

[1998] See the Harassment, stalking, threats and abuse and Controlling or coercive behaviour chapters

[1999] Law Commission, 2021. Modernising Communications Offences: A final report. [accessed 26 July 2021].

[2000] Q21 (Table 1527). Source: Ofcom, 2023. Experiences of using online services. [. [accessed 1 August 2023].

[2001] Survey was carried out on OnePoll and the sample consisted of 4,000 single respondents aged 18-65 who were actively using dating sites. Source: Plenty of Fish, 2023. The Desirable Dating Guide. [accessed 28 July 2023].

[2002] McGlynn, C. and Johnson, K., 2022.

Risk factors: User base

User base size

19.26 There is no evidence to indicate that user base size is a specific risk factor for the cyberflashing offence. However, we expect the number of users on a service could play a role in a similar manner to that presented in the introduction's 'Understanding illegal harms online' section.

User base demographics

19.27 The following section outlines key evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

19.28 Data suggests that user base characteristics including gender, age, race and ethnicity, and sexual orientation could lead to an increased risks of harm to individuals.

19.29 Cyberflashing is a gendered harm; among individuals aged 18-34, women are much more likely than men to have received an unsolicited sexual photo (40% vs 26%).[2003]

19.30 The evidence suggests that cyberflashing is most prevalent among young people; 47%-48% of women aged 18-24 have been cyberflashed. [2004] [2005]

19.31 There is some evidence that cyberflashing is more prevalent as a harm for minority ethnic groups. Internet users in minority ethic groups were more likely than white internet users to have seen or experienced cyberflashing in the past four weeks (8% vs 3%).[2006]

19.32 There is some evidence that cyberflashing is more prevalent as a harm for LGBT+ people. Internet users who identify as gay, lesbian (5%) or bisexual (7%) were more likely than internet users who identify as heterosexual (3%) to have experienced cyberflashing in the past four weeks.[2007]

Risk factors: Functionalities and recommender systems

User identification

User profiles

19.33 The ability to create multiple user profiles is a risk factor for cyberflashing offences. Perpetrators can create profiles which represent particular identities from which they can target their victims and cyberflash. If reported by the victim, or blocked, perpetrators have

[2003] Sample consisted of 1629 adults. Source: YouGov (Smith, M), 2018. Four in ten young women have been sent unsolicited sexual images. [accessed 28 July 2023].

[2004] Survey commissioned by Bumble and was carried out by Research without Borders, spanning between 15th-18th October 2021 with 1,793 respondents who live in England or Wales: Bumble, n.d. Women's Safety Experts Join Bumble to Call on UK Prime Ministerial Candidates to Prioritise Anti-Cyberflashing Law. [accessed 10 June 2024]

[2005] Sample consisted of 2,353 women aged 18-36 and 1,327 men aged 18-36. Source: YouGov (Smith, M), 2018. Four in ten female millennials have been sent an unsolicited penis photo. [accessed 28 July 2023].

[2006] Ofcom, Online Experiences Tracker 2021-2022. Comprises Wave 1 and 2 combined data set.

[2007] Ofcom, Online Experiences Tracker 2021-2022. Comprises Wave 1 and 2 combined data set.

been able to create another account and an associated user profile to continue their offence.[2008]

Anonymous user profiles

19.34 The ability to create anonymous user profiles could be a risk factor for cyberflashing offences. Research has found that anonymity creates a disinhibition effect[2009] which could lead perpetrators to engage in harmful behaviour such as cyberflashing.

User networking

User connections and user searching

19.35 Functionalities that allow users to find and contact one another, such as user connections and user searching, are a risk factor for cyberflashing offences. These functionalities allow potential perpetrators to make contact with victims and survivors who were previously unknown to them in order to cyberflash.[2010]

19.36 Victims and survivors describe unknown perpetrators making contact with them on social media in order to cyberflash.[2011] Several high-profile women have also shared their experiences of being cyberflashed by men who made contact with them on social media.[2012]

User communication

Direct messaging

19.37 Direct messaging is a risk factor for cyberflashing offences. Victims and survivors describe perpetrators sending messages which contain unsolicited sexual images.[2013] Sixty-six per cent of women who have experienced cyberflashing and 59% of men were using a text or messaging app at the time.[2014]

Ephemeral messaging

19.38 Some U2U services allow users to send ephemeral messages, which in some cases disappear as soon as they have been seen by the receiver. There is evidence that ephemeral messaging is a risk factor for cyberflashing offences as it allows perpetrators to send non-

---

[2008] Center for Countering Digital Hate, 2022. Hidden Hate: How Instagram fails to act on 9 in 10 reports of misogyny in DMs. [accessed 28 July 2023]; Bond, K., 2023. 'It's in our phone, in our hands, and it's in our house': Six women share their experiences of online sexual harassment, *Metro,* 31 January. [accessed 28 July 2023].

[2009] Suler, J., 2004. The Online Disinhibition Effect, *Cyberpsychology & behaviour: the impact of the internet, multimedia and virtual reality on behaviour and society,* 7(3). [accessed 21 September 2023].

[2010] McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. Bristol: Bristol University Press. [accessed 21 September 2023].

[2011] Gallagher, S., 2018. 'Violated, Sick, Uncomfortable': 10 Women On Being Sent Unsolicited Dick Pics, *HuffPost,* 26 October. [accessed 28 July 2023]; Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 28 July 2023].

[2012] BBC, 2023. Emily Atack: Asking For It? (Documentary), 31 January. [accessed 2 August 2023]; Kelly, K., 2023. Love Island star Amy Hart felt 'violated' after being bombarded with 'cyberflashing' online, *LBC,* 18 April. [accessed 28 July 2023]; McLoughlin, L., 2023. Vanessa Feltz reveals she 'regularly' receives unsolicited sexual images from men, *Evening Standard,* 19 April. [accessed 28 July 2023].

[2013] McGlynn, C. and Johnson, K., 2022; Revealing Reality, 2022.

[2014] Sample consisted of 1,629 adults. Source: YouGov (Smith, M), 2018. Four in ten young women have been sent unsolicited sexual images. [accessed 28 July 2023].

permanent images. Young people in particular report being cyberflashed via ephemeral messaging services.[2015]

Posting content (images and videos)

19.39 The ability to post content, in this case images or videos, is a risk factor for cyberflashing offences. It is a functionality which enables perpetrators to upload sexual images and cyberflash other users.[2016]

Livestreaming

19.40 There is some evidence that the ability to livestream is a risk factor for cyberflashing offences. Sometimes known as 'zoomflashing', this form of cyberflashing occurs when a perpetrator exposes themselves in real time through a video-call system.[2017]

Content storage and capture

Capturing images

19.41 U2U services which use a device's camera hardware to create images (using filters, for example) can also be considered a risk factor likely exploited by perpetrators to create and share images easily on a service.

Risk factors: Business models and commercial profile

19.42 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[2015] Ringrose, J., Regehr, K. and Whitehead, S, 2021. Teen Girl' Experiences Negotiating the Ubiquitous Dick Pic: Sexual Double Standards and the Normalization of Image Based Sexual Harassment, *Sex Roles, 85* (558). [accessed 11 August 2023]; Revealing Reality, 2022.

[2016] Revealing Reality, 2023. Anti-social Media: The violent, sexual and illegal content children are viewing on one of their most popular apps. [accessed 26 July 2023].

[2017] Elmer, G., Neville, S., Burton, A. and Ward-Kimola, S., 2021. Zoombombing During a Global Pandemic, *Social Media + Society,* 7(3). [accessed 28 July 2023].

20. Encouraging or assisting serious self-harm

Warning: This chapter contains discussion of suicide, self-harm and eating disorders.[2018]

Summary analysis for encouraging or assisting suicide (or attempted suicide) or serious self-harm offences: how harm manifests online, and risk factors

This offence takes place when an individual intentionally encourages or assists a person to carry out serious self-harm. Ofcom's Online Experiences Tracker showed that 4% of UK internet users had seen or experienced content 'promoting self-harm' in the past four weeks preceding the survey.[2019] Younger respondents were more likely to see or experience this content, with 7% of 13-to-24-year-olds and 25-to-34-year-olds, and 10% of 18-to-24-year-olds, compared to 3% or fewer of users in age groups over 35.

The physical and psychological harms that can arise from these offences are severe and can include long-term mental health concerns, eating disorders, physical harm to oneself, and death. Harm from these offences can affect both viewers of the content and the user posting the content themselves.

Posting of content that amounts to the offense of encouraging or assisting serious self-harm is often found in proximity[2020] to posting of other types of content related to self-harm that do not amount to the offense. While most content related to self-harm does not amount to the offense, all content related to self-harm is extremely sensitive and may have the potential to cause harm to users. While there may be users who post self-harm related content to cause harm to others, some users may share this content to find supportive communities, to express their own experiences as part of a healing process or to attempt to help others. Users posting and engaging with this type of content can include those in vulnerable circumstances who are themselves dealing with thoughts of suicide or self-harm, as well as those who have recovered or are recovering from mental health challenges.

There are ethical and legal limitations to conducting research into this type of content, and research has often relied on correlational or qualitative methods for insights into risk factors.

Service type risk factors:

[2018] If you need support, please check the following websites: NHS, 2023. Help for suicidal thoughts.; NHS, 2023. Where to get help for self-harm; Beat, 2023. Helplines for eating disorder support.
[2019] Ofcom, 2024. Online Experiences Tracker. [accessed 22 November 2024]. Fieldwork was carried out in a four-week period in May and June 2024. Content users reported seeing may include content that could be deemed illegal.
[2020] For example using the same hashtags, posted to the same accounts, or in the same user groups.

Discussion forum and chat room services can act as spaces where self-harm is assisted or encouraged. They may be exploited by individuals with the intent to cause harm or distress among users who are experiencing thoughts of self-harm. This may be particularly true of services that facilitate discussions on niche or specialised content among smaller groups of users, which could include suicide or self-harm content. However, discussion forums and chat rooms may also be used by individuals experiencing mental health difficulties to connect with other users for support and guidance. **Social media services** can allow potential perpetrators to disseminate self-harm related content. Users can view self-harm content on these types of services, particularly through the creation of user groups on social media services.

Services that allow users to build **online communities** are also a risk factor, as online communities can act as spaces where eating disorders are promoted or encouraged.

User base risk factors:

Small and large **user base sizes** can pose risks, for different reasons. With a larger user base, more people risk encountering this content, while smaller user bases can foster the sharing of specialised and extreme content relating to self-harm and eating disorders.

Users who are in vulnerable circumstances such as those suffering with **their mental health** and who might be experiencing thoughts of suicide or self-harm are more likely than other users to be at risk from this type of content.

Age is also a potential risk factor. Our evidence suggests that young people are more likely to encounter this content, to use the internet for self-harm related purposes, and to be susceptible to copycat behaviour.

Functionalities and recommender systems risk factors:

Commenting on content is a risk factor, there is evidence of people using comments to encourage the self-harm of the person that distributed the content. Commenting on content intersects with other risk factors such as **livestreaming and posting content** to create a high-risk context. For example, livestreaming is a risk factor that has been used to share real-time acts of self-harm. While livestreaming these activities is not in itself illegal, the social functionality attached to the livestream, including commenting on the livestream or in user groups connected to the livestream can be used to encourage the self-harm depicted on the livestream. Similarly, comments on posts related to self-harm can encourage the user, who may be experiencing thoughts of self-harm, to attempt to take their own life.

Anonymous user profiles appear to be a risk factor. Some users may feel more confident in sharing content depicting or discussing harmful themes if they cannot be identified, including assisting or inciting others in acts of suicide or serious self-harm. However, users may also feel that anonymity allows them to talk more

openly about their own thoughts of suicide and self-harm, and to connect with others with similar experiences. In this context, anonymity can both pose risks and confer potential benefits to those seeking help.

The **ability to post content** and **re-post or forward content** can allow users to connect with others who are experiencing similar thoughts or behaviours in a beneficial way, but it can also be used to disseminate harmful self-harm content.

Content recommender systems can also be a risk factor for this type of content. The way in which recommender systems are designed can influence the extent to which harmful (and potentially illegal) content is recommended to users. Recommender systems are commonly designed to optimise for user engagement, and learn about users' preferences through implicit (e.g., viewing multiple times) and explicit (e.g., liking, sharing, and commenting) user feedback. While further evidence is needed, research suggests that where there are vulnerable users who are engaging with harmful content, such as self-harm content, recommender systems are more likely to create a 'filter bubble' or 'rabbit hole.' This may lead to users discovering more content that is harmful or distressing, as well as potentially illegal. If a user is primarily engaging with harmful content, then this is likely to create a filter bubble where the user is recommended more harmful content, while other content is deprioritised.

Other functionalities risk propagating this offence. **Content tagging** such as hashtags can help evade content moderation techniques on self-harm content, because groups of users create hashtags that differ from those that may be blocked as harmful. **Group messaging** can also enable users to contact one another and can encourage harmful behaviour in a group setting.

Introduction

20.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on user-to-user (U2U) services that may amount to the offence of encouraging or assisting serious self-harm detailed under 'Relevant offences' below; and
- the use of these services for the commission and/or facilitation of this offence (collectively the 'risks of harm').

20.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly

encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.[2021]

20.3 Serious self-harm is not a criminal offence; nor is discussing or portraying serious self-harm in any way which does not amount to encouraging or assisting an illegal act. An individual attempting to do this will not and should not be penalised. Individuals suffering with their mental health or in otherwise vulnerable circumstances, including those dealing with thoughts of self-harm, should be able to seek support without fear of negative consequences.

20.4 Services should be aware that content relating to self-harm varies and is not always shared with intent to encourage or assist self-harm. Users who share self-harm content may themselves be vulnerable and are using online spaces to express their feelings and seek support by connecting with others who may be having similar experiences.[2022] Services should therefore be mindful of this distinction when assessing this type of content, considering the risks of harm to the user who shares the content as well as to other users.

20.5 In this chapter we explore the evidence related to an increased risk of encountering content or activity online that may amount to the offence of encouraging or assisting serious self-harm. In some cases we have considered evidence that covers both self-harm and suicide-related content, as it is often difficult to distinguish between content that focuses solely on suicide, compared to content which focuses on self-harm that is potentially life-threatening.[2023] For the purposes of our assessment, including to assess the impact of characteristics on the risks of harm, we also treat some self-harm content as *potentially* amounting to illegal content, although we recognise that whether it is illegal content depends on the intentions of the person sharing the content (see 'Relevant offences' and the ICJG for details on how to identify illegal content).

20.6 Where data from Ofcom's Online Experience Tracker is included, this is based on participants' self-reported experience of having seen or experienced 'content relating to self-harm or suicide', which may not necessarily include content deemed to meet the illegal threshold.

Relevant offences

20.7 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. In regard to self-harm, Ofcom is required to consider the risks of harm

[2021] As with other chapters, we have considered evidence of self-harm content from a variety of sources, including information provided by services, academic literature, third-party research and civil society in general. Some of this evidence relates to content which may not necessarily mirror, or is broader than, the criminal definitions of this offence, and is closely related to content that encourages or assists self-harm.

[2022] For example, a UK-based qualitative study with participants who had either previously used the internet for suicide-related purposes, or had been admitted to hospital following serious self-harm, found that a number of young adults in the sample used discussion forums and chatrooms, information-sharing services (in this case, Q&A websites), and social media services to express their feelings, manage loneliness or engage in dialogue with others. Source: Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[2023] Brennan, C., Saraiva, S., Mitchell, E, Melia, R., Campbell, L., King, N. and House, A., 2022. Self-harm and suicidal content online, harmful or helpful? A systematic review of the recent evidence, *Journal of Public Mental Health,* 21 (1). [accessed 10 July 2023].

connected with the non-priority communications offence listed in Part 10, Section 184 of the Act.

20.8 In this chapter, we consider the non-priority communications offence of encouraging or assisting serious self-harm.[2024]

20.9 The offence of assisting or encouraging serious self-harm is committed where a person does an act capable of encouraging or assisting the serious self-harm of another person. 'Serious self-harm' is defined as self-harm that would amount to grievous bodily harm.[2025] This offence also requires the user posting to have had intent to encourage or assist.

20.10 For this offence, it is not necessary for the encouragement or assistance to be targeted towards a specific person or persons. The content does not need to result in actual self-harm to amount to illegal content.

20.11 For more details on the offences and how services can assess whether content amounts to illegal content, please refer to the Illegal Content Judgements Guidance or ICJG.

How encouraging or assisting serious self-harm manifests online

20.12 This section is an overview which looks at how the offence of encouraging or assisting serious self-harm manifests online, and how users may be at risk of harm.

20.13 To put the risks of harm from this offence into context, Ofcom's 2024 Online Experiences Tracker found that 4% of UK internet users had seen or experienced content 'promoting self-harm' in the past four weeks.[2026] Younger respondents were more likely to see or experience this content, with 7% of 13-to-24-year-olds and 25-to-34-year-olds, and 10% of 18-to-24-year-olds, compared to 3% or fewer of users in age groups over 35.

20.14 Self-harm content can manifest online in various forms, with a range of effects on individuals. Samaritans, a charity that works with people struggling to cope and people at risk of suicide, notes examples of suicide or self-harm content that may pose a risk to individuals.[2027] [2028] These include detailed and instructive information about suicide or serious self-harm methods, posts encouraging, glamourising or celebrating suicide or serious self-harm, and graphic images relating to serious self-harm or suicide.[2029] Not all of this type of content will be illegal, though all of it has the potential to be harmful.

[2024] Section 184 of the Online Safety Act 2023.
[2025] The Illegal Content Judgements Guidance indicates that a specific kind of eating disorder content could be considered illegal under the spirit of the offence. To that end, we have included evidence on the risks of harm related to types of eating disorder content in this chapter. We note that many risk factors are used in a similar way as suicide and self-harm content.
[2026] Ofcom, 2024. Online Experiences Tracker. [accessed 22 November 2024]. Fieldwork was carried out in a four-week period in May and June 2024. Content users reported seeing may include content that could be deemed illegal.
[2027] Samaritans, 2022. Towards a suicide-safer internet. [accessed 24 May 2023]
[2028] These examples may not necessarily be illegal in all cases. For information on what could be considered illegal, please refer to the ICJG guidance.
[2029] The full list of examples includes: detailed and instructive information about suicide or self-harm methods; posts encouraging, glamourising or celebrating suicide or self-harm; posts from people seeking or encouraging suicide or self-harm pacts; posts relating to suicide or self-harm challenges; graphic images relating to self-harm or suicide; and livestreams or recorded videos of suicidal or self-harming behaviour.

20.15 While it is challenging to quantify the social and economic cost of this harm, researchers from the University of Oxford and the London School of Economics estimated that the average cost for each episode of self-harm recorded between April 1, 2013 and March 31, 2014 was £809 (£1064.11 in January 2024 prices).[2030]

Risks of harm to individuals presented by the offences of encouraging or assisting serious self-harm online

20.16 There are at least two distinct groups of users who are likely to be at risk: those who encounter this content unintentionally (e.g. when searching for content that overlaps with hashtags to share harmful self-harm content, or when served up by a recommender system), and those who may be experiencing thoughts of suicide or serious self-harm and are seeking this type of content. Other users at risk may include those who are looking to disengage from suicide or self-harm content but encounter this content again, having previously engaged with suicide or self-harm content.

20.17 Several studies have further explored the impact of exposure to suicide and self-harm related content. A study of 18-to-29-year-olds found that those who viewed content depicting self-harm on a social media service, either intentionally or by accident, were at a higher risk of self-harm and suicide.[2031] The potential negative effects of this content were also evident in a national survey by Swansea University and Samaritans (where 87% of the sample[2032] reported having self-harmed before). It asked respondents about the impact of seeing or sharing self-harm or suicide content online: 35% reported a worsening of their mood, with only 2% reporting that this type of content improved their mood.[2033] However, more than half the respondents reported that the impact this content had on them depended on their mood at the time, so the proportion whose mood was negatively affected is potentially higher than 35%. A recent review of 15 studies on the potential impacts of viewing self-harm related images online found both harmful and protective effects. All 15 studies found evidence of harmful effects. These included being 'triggered' by the images, normalising or escalating self-harm through sharing tips and ideas and being encouraged to share images or compete with others.[2034] Nine of the 15 studies found evidence of protective effects.[2035] Reporting both protective and harmful experiences simultaneously is not uncommon in the research. For example, in research exploring young

[2030] Costs are mainly driven by healthcare services used, including intensive care and psychosocial assessment. Therefore estimated cost of each episode varies substantially across types of self-harm. Source: Tsiachristas, A., Geulayov, G., Casey, D., Ness, J., Waters, K., Clements, C., Kapur, N., McDaid, D., Brand, F., & Hawton, K. (2020). Incidence and general hospital costs of self-harm across England: estimates based on the multicentre study of self-harm. *Epidemiology and psychiatric sciences*, 29. [accessed 10 October 2024]

[2031] The study found that exposure to depictions of self-harm on Instagram resulted in an 'emotional disturbance' in some users, with this exposure statistically positively associated with psychometric predictors of "(possibly harmful) self-harm and suicidality-related outcomes". Source: Arendt, F., Scherr, S. and Romer, D., 2019. Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults, *New Media & Society,* 21 (11-12). [accessed 10 July 2023].

[2032] The sample included 5,294 individuals aged 16-84 years. Many of the participants in the study were females aged under 25, and so does not represent any population as a whole. Source: Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. [accessed 10 July 2023].

[2033] Samaritans and Swansea University, 2022.

[2034] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023. Research Review: Viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms, *Journal of Child Psychology and Psychiatry,* 64 (8). [accessed 10 July 2023].

[2035] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023.

people's experience with mental health difficulties, one 17-year-old girl described how viewing suicide and self-harm related content online kept her calm and occupied, while at the same time desensitised her to the act of self-harm and reinforced self-harm objectives.[2036]

20.18 It is likely that repeated exposure to self-harm-related content within online communities normalises the act of self-harm.[2037] Furthermore, self-harm content may have a 'contagion' effect whereby being exposed to the idea of self-harm or graphic images of self-harm increases the risk that someone attempts to undertake a similar act. The Royal College of Psychiatrists note that "access to graphic images and the normalisation of self-harm through online forums is of significant concern and reflects the experience of clinicians, for example, regarding the well-known 'contagion' effects of self-harm in inpatient units".[2038]

20.19 According to Samaritans, evidence suggests that content presenting suicide or self-harm behaviours (such as viral suicide and serious self-harm 'challenges', encouraging users to engage in harmful behaviour), may encourage or assist other users to undertake acts of suicide and self-harm. They state that the contagion effect may become more likely, increasing the risk of imitation, when the viewer overly identifies with the original uploader of the content (for example, if they are at increased risk of thoughts of suicide or serious self-harm).[2039]

20.20 In this context of increased access to the internet, online services and potentially harmful content over the past decades, it is notable that non-suicidal self-injury (NSSI) hospital admission rates among 10-to-24-year-olds in the UK increased 23% from 348.9 per 100,000 in 2012/13 to 427.3 per 100,000 in 2021/22.[2040] Rates of NSSI admissions for 10-to-14-year-olds and 15-to-19-year-olds, have increased substantially.[2041]

20.21 Content that encourages eating disorders, which can carry substantial risks of harm, could potentially be linked to this offence.[2042] Eating disorders are associated with adverse physical consequences and suicide risk, and anorexia nervosa has the highest mortality rate of any mental illness.[2043] A recent survey of 255 people with lived experience of eating disorders run by Beat, the UK's eating disorder charity, found that 91% of respondents had

---

[2036] Livingstone, S., Stoilova, M., Stoilova, M., Stänicke, L. I., Jessen, R. S., Graham, R., Staksrud, E., Jensen, T. 2022. Young people experiencing internet-related mental health difficulties: the benefits and risks of digital skills, *ySKILLS*. [accessed 10 October 2024].

[2037] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023.

[2038] Royal College of Psychiatrists, 2020. Technology use and the mental health of children and young people. [accessed 10th July 2023].

[2039] Samaritans, 2022. Towards a suicide-safer internet. [accessed 24 May 2023]

[2040] Please note that the figure for 2022/23 is much lower at 319/100,000, but the Nuffield Trust attributes this to "(…) the change in NHS England's reporting methodology, which reclassified Same Day Emergency Care (SDEC) cases, leading to fewer admissions being recorded under the Admitted Patient Care data set." Source: Nuffield Trust, 2024. Hospital admissions as a result of self-harm in children and young people. [accessed 1 October 2024]

[2041] NSSI admissions of patients aged 10-14-years-old increased 148% to 307.1 per 100,000 in 2021/22 from 123.8 in 2012/13. NSSI admissions of patients aged 15-19-years-old increased 37% to 641.7 per 100,000 in 2021/22 from 469.2 in 2012/13: Nuffield Trust, 2024.

[2042] Feeding and eating disorders, as defined by the ICD-11, include anorexia nervosa, bulimia nervosa, binge eating disorder, other specified food intake disorder (OSFED), avoidant restrictive food intake disorder (ARFID), rumination disorder and pica. Source: ICD, 2023. Feeding or eating disorders. [accessed 10 July 2023].

[2043] Smith, A., Zuromski, K. and Dodd, R., 2018. Eating disorders and suicidality: what we know, what we don't know, and suggestions for future research, *Current Opinion in Psychology,* 22. [accessed 10 July 2023].

encountered content which they described as harmful in the context of their eating disorder.[2044]

20.22 Exposure to pro-eating disorder (pro-ED) content has been shown to be associated with increases in eating disorder behaviours, and can play an important role in the causes of eating disorders.[2045] Use of pro-ED websites has also been shown to discourage help-seeking and maintain or prolong an eating disorder.[2046] As early intervention is key to effective treatment for an eating disorder, delays to help-seeking and treatment are significant.[2047] [2048]

20.23 Some evidence suggests that 13-to-17-year-old females are the demographic group most likely to visit pro-eating disorder websites, but although this is not considered in depth in this chapter due to its focus on adults, it is important to note that eating disorders can affect anyone, of any gender or age.[2049] [2050] Indeed, it has been estimated that 25% of people with an eating disorder are male, and a recent report by the Eating Disorder Genetics Initiative found that eating disorders are just as likely to start in adulthood as in childhood.[2051] [2052] [2053]

20.24 Encouraging or assisting eating disorders can take many forms online. For instance, users can pose as online 'coaches', targeting young people with a technique called 'meanspo' (being mean in order to inspire or encourage eating disorder behaviours). Examples include 'coaches' verbally abusing young people on social media services or through direct messages to encourage their eating disorders. This was highlighted in the BBC Three documentary *Zara McDermott: Disordered Eating.*[2054] Users will also occasionally request 'meanspo' from other users, as a form of inspiration and enforcement for their disordered eating.[2055] Children with experience of an eating disorder may also be more vulnerable to

---

[2044] Beat, 2023. Online Safety and Eating Disorders, [accessed 10 July 2023].

[2045] Mento, C., Silvestri, M C., Muscatello, M R A., Rizzo, A., Celebre, L., Praticò, M, Zoccali, R A. and Bruno, A., 2021. Psychological Impact of Pro-Anorexia and Pro-Eating Disorder Websites on Adolescent Females: A Systematic Review. *International journal of environmental research and public health,* 18 (4). [accessed 10 July 2023].

[2046] Gale, L., Channon, S., Larner, M. and James, D., 2015. Experiences of using pro-eating disorder websites: a qualitative study with service users in NHS eating disorder services, *Eating and Weight Disorders – Studies on Anorexia, Bulimia and Obesity,* 21. [accessed 10 July 2023]

[2047] Beat, 2022. Best practice in ensuring early intervention for eating disorders. [accessed 10 July 2023].

[2048] Other research demonstrates that the first three years of an eating disorder are a critical window after which symptoms can become more entrenched. Source: Treasure, J., Stein, D. and Maguire, S., 2015. Has the time come for a staging model to map the course of eating disorders from high risk to severe enduring illness? An examination of the evidence, *Early Intervention in Psychiatry,* 9 (3). [accessed 10 July 2023].

[2049] Ofcom's risk assessment for content harmful for children, coming out in spring 2024, will consider content that encourages, promotes or provides instructions for an eating disorder or behaviours associated with an eating disorder.

[2050] Mento, C., Silvestri, M C., Muscatello, M R A., Rizzo, A., Celebre, L., Praticò, M, Zoccali, R A. and Bruno, A., 2021. Psychological Impact of Pro-Anorexia and Pro-Eating Disorder Websites on Adolescent Females: A Systematic Review. *International journal of environmental research and public health,* 18 (4). [accessed 10 July 2023].

[2051] Sweeting, H., Walker, L., MacLean, A., Patterson, C., Räisänen, U and Hunt, K., 2015. Prevalence of eating disorders in males: a review of rates reported in academic research and UK mass media, *International Journal of Mens Health,* 14 (2). [accessed 11 July 2023].

[2052] Wooldridge, T., Mok, C. and Chiu, S., 2014. Content analysis of male participation in pro-eating disorder web sites, *Eating Disorders,* 22 (2). [accessed 11 July 2023].

[2053] King's College London and Beat (Davies, H L., Kelly, J., Ayton, A., Hübel, C., Bryant-Waugh, R., Treasure, J. and Breen, G)., 2022. When Do Eating Disorders Start? An Investigation into Two Large UK Samples. [accessed 11 July 2023].

[2054] Wales Online, 2022. Coaches are training children to be anorexic with vile comments online. [accessed 10 July 2023].

[2055] Achilles, L., Mandl, T. and Womser-Hacker, C., 2022. "Meanspo Please, I Want to Lose Weight": A Characterization Study of Meanspiration Content on Tumblr Based on Images and Texts. [accessed 10 July 2023].

online grooming, as there is evidence to suggest that some perpetrators deliberately target these children. We recommend readers refer to the Grooming chapter of the Register of Risks for a complete understanding of this issue and how it relates to, and can constitute, grooming offences.

Evidence of risk factors on user-to-user services

Risk factors: Service types

20.25 Research indicates that the following types of services can be used to commit or facilitate the offence of encouraging or assisting serious self-harm: discussion forums and chat rooms, information-sharing services, social media services, video-sharing services and services that more generally enable community building.

Discussion forums and chat rooms, information sharing services

20.26 Our evidence shows that discussion forums and chat room services can act as spaces where self-harm is assisted or encouraged. Although these services can have positive benefits, they can also facilitate discussion and ideation relating to self-harm, which can escalate into encouragement of suicidal behaviours, including sharing content that can be harmful or distressing to users.[2056] Users may also specifically post to prompt other users to provide them with information regarding how to carry out serious self-harm.

20.27 A number of deaths in the UK in recent years have reportedly involved chatrooms and forums,[2057] and the Royal College of Psychiatrists has emphasised the normalisation of sharing graphic images of self-harm on discussion forums as a significant concern among those who are already vulnerable.[2058]

20.28 Services that enable users to build online communities are a risk factor. Research conducted in August 2022 by the Network Contagion Research Institute found a growing community of individuals mutually promoting, celebrating and encouraging self-harm on Twitter (now 'X'). The community uses a coded language within hashtags and account biographies to evade keyword-based moderation algorithms.[2059] This language includes terms referring to different levels and types of self-injury. Posts within this community receive high levels of engagement. For example, one post depicting open self-inflicted wounds received 2000 likes, 165 retweets and 80 comments. User comments on the image included problematic reinforcement, for example "that's so pretty", "how beautiful", and "what did you use."[2060] Likewise for eating disorders, a 2021 study was able to demonstrate

[2056] There are services such as some discussion forums that are dedicated to suicide or self-harm content. However, a recent inquest has revealed that this content also exists across services that actively prohibit suicide or self-harm content. Source: The Coroner's Service, 2022. Prevention of Future Deaths. [accessed 28 October 2022].
[2057] In the UK between 2001 and 2008, there were at least 17 deaths involving chatrooms or sites that provide advice on suicide methods. Source: Cohen-Almagor, R, and Lehman-Wilzig, S., 2022. Digital Promotion of Suicide: A Platform-Level Ethical Analysis, *Journal of Media Ethics,* 32 (2). [accessed 10 July 2023].
[2058] Royal College of Psychiatrists, 2020. Technology use and the mental health of children and young people. [accessed 10 July 2023]. *See* Risk of harm to individuals presented by the offences of encouraging or assisting suicide or self-harm online *for more information.*
[2059] This is sometimes referred to as 'algospeak'.
[2060] [DISTRESSING CONTENT WARNING] Network Contagion Research Institute, 2022. Online Communities of Adolescents and Young Adults Celebrating, Glorifying, and Encouraging Self-Harm and Suicide are Growing Rapidly on Twitter. [accessed 10 October 2024]

sustained decreases in the body mass index (BMI) of users who joined a pro-ED online community, with the more active users losing more weight during the period they were involved.[2061] Similar results have been found in individuals with no history of an ED.[2062]

20.29 While pro-ED communities have been argued to offer a sense of emotional support, this support can depend on unhealthy group conformity.[2063] Pro-ED content can also involve images to inspire weight loss (so called 'thinspiration' or 'thinspo').[2064] Since there is some evidence to suggest that competitiveness can be associated with eating disorder pathology, the social aspects of pro-ED communities can be particularly harmful for people with, or vulnerable to, eating disorders.[2065]

Social media services

20.30 The available evidence suggests that social media services can play a role in the dissemination of harmful self-harm related content. Research shows that users who view self-harm on social media services,[2066] either intentionally or by accident, are at a higher risk of self-harm and suicide.[2067] Dedicated self-harm or suicide groups are also occasionally set up by users on social media services, offering users a chance to discuss topics with other users.[2068]

20.31 Our evidence indicates that social media services that focus on allowing users to share images may be more closely linked to body dissatisfaction than those that are not.[2069]

Video-sharing services

20.32 Video-sharing services can also play a role in disseminating suicide and self-harm content. There have been several cases in which livestreaming, a functionality common to video-sharing services, has been used to show users self-harming or ending their life in real

---

[2061] Feldhege, J., Moessner, M. and Bauer, S., 2021. Detrimental Effects of Online Pro-Eating Disorder Communities on Weight Loss and Desired Weight: Longitudinal Observational Study, *Journal of Medical Internet Research,* 23 (10). [accessed 11 July 2023].

[2062] In a US study, adult females without history of an ED who were exposed to pro-eating disorder content for 90 minutes showed a significant decrease in their calorific intake from pre- to post-exposure, had strong emotional responses to the content and reported changes in their current eating behaviour three weeks after the study. Source: Jett, S., LaPorte, D J. and Wanchisn, J., 2010. *Impact of exposure to pro-eating disorder websites on eating behaviour in college women*, *European Eating Disorders Review,* 18 (5). [accessed 11 July 2023].

[2063] Feldhege, J., Moessner, M. and Bauer, S., 2021. Detrimental Effects of Online Pro-Eating Disorder Communities on Weight Loss and Desired Weight: Longitudinal Observational Study, *Journal of Medical Internet Research,* 23 (10). [accessed 11 July 2023].

[2064] Ging, D. and Garvey, S., 2018. 'Written in these scars are the stories I can't explain': A content analysis of pro-ana and thinspiration image sharing on Instagram, *New Media and Society,* 20 (3). [accessed 11 July 2023].

[2065] Osborne,K D., 2023. Competing for perfection: a scoping review evaluating relationships between competitiveness and eating disorders or disordered eating behaviours. [accessed 8 September 2023].

[2066] In particular, Instagram, as the focus of this study at the time.

[2067] The study found that exposure to this content resulted in an 'emotional disturbance' in some users, with this exposure statistically related to '(possibly harmful) self-harm and suicidality-related outcomes'. Source: Arendt, F., Scherr, S. and Romer, D., 2019. Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults, *New Media & Society,* 21 (11-12). [accessed 10 July 2023].

[2068] Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., Purdy, N., Daine, K. and John, A., 2017. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown, *PLoS ONE,* 12 (8). [accessed 10 July 2023]. See Risk factors: functionalities and recommender systems section for more information.

[2069] Harriger, J A., Evans, J A., Thompson, J K. and Tylka, T L., 2022. The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms, *Body Image,* 41. [accessed 11 July 2023].

time.[2070] These livestreams occasionally also contain potentially illegal content within the comment threads (see Commenting on content for more information).

Risk factors: User base

User base size

20.33 Services with both large and small user bases pose risks in relation to self-harm content, for different reasons.

20.34 On the one hand, the larger a service's user base, the greater the number of people who are likely to encounter content on it, particularly where it is amplified through recommender systems, meaning that content can receive substantial amounts of engagement.[2071] This in turn heightens the risk of contagion effects, as described earlier.

20.35 Meanwhile, services with a small user base may be more likely to foster the sharing of more niche or specialised content, which could include suicide or self-harm content.

20.36 Pro-ED content, for example, can appear on large services as well as on smaller services, websites and blogs.[2072] [2073] Use of specialised pro-ED websites can expose users to extreme content and present a barrier to eating disorder recovery.[2074] [2075]

User base demographics

20.37 The following section outlines the key evidence of user base demographic factors and risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

20.38 The data suggests that user base characteristics including **age, mental health, ethnicity, sexual orientation,** and **gender** could lead to increased risks of harm to individuals.

20.39 The age of users appears to influence their likelihood of encountering content related to self-harm. Online Experiences Tracker (OET) data shows that younger respondents were slightly more likely to see or experience this content, with 7% of 13-to-24-year-olds, and 10% of 18-to-24-year-olds – for age groups over 35, at most 3% of respondents reported seeing this type of content.[2076] The tendency for younger users to encounter self-harm related content online more frequently may be corroborated by Instagram's own Bad Experiences and Encounters Framework findings from 2021. Younger Instagram users were

[2070] For an explanation, see 'livestreaming' in Risk factors: functionalities and recommender systems section.
[2071] Ekō, 2023. Suicide, Incels, and Drugs: How TikTok's deadly algorithm harms kids. [accessed 11 July 2023].
[2072] Feldhege, J., Moessner, M. and Bauer, S., 2021. Detrimental Effects of Online Pro-Eating Disorder Communities on Weight Loss and Desired Weight: Longitudinal Observational Study, *Journal of Medical Internet Research,* 23 (10). [accessed 11 July 2023].
[2073] Smahelova, M., Drtilova, H., Smahel, D., Cevelicek, M, 2020. Internet Usage by Women with Eating Disorders during Illness and Recovery, *Health Communication* 35 (5). [accessed 24 July 2023].
[2074] Gale L, Channon S, Larner M, James D. 2016 Experiences of using pro-eating disorder websites: a qualitative study with service users in NHS eating disorder services. *Eating and Weight Disorders,* 21(3). [accessed 24 July 2023].
[2075] Rodgers, R.F., Melioli, T, 2016. The Relationship Between Body Image Concerns, Eating Disorders and Internet Use, Part I: A Review of Empirical Support. *Adolescent Res Rev* 1, 95–119 [accessed 24 July 2023].
[2076] Ofcom, 2024. Online Experiences Tracker 2024.

more likely to report "*see[ing] someone harm themselves, or threaten to do so, on Instagram*", especially those aged 13-15 (8.4%).[2077]

20.40 Evidence also suggests that younger adults are more likely to experience the contagion effect (i.e., copycat behaviour)[2078] and to have used the internet for suicide-related purposes (among those who had been in contact with mental health services).[2079]

20.41 A study by Samaritans and Swansea University found that people with a history of self-harm were more likely to report that they were 10 years old or younger when they first viewed self-harm or suicide content online, whereas those with no history of self-harm were more likely to report being aged 25+ at the time of first encountering this content.[2080] In response to an Ofcom call for evidence, Samaritans suggested that this may indicate a potential correlation between viewing harmful content at a young age and future harmful behaviour.[2081]

20.42 As noted previously, this increased likelihood to be exposed to this content comes at a time when rates of admission to hospital for non-suicidal self-injury have increased dramatically among young people.

20.43 Neurological and psychological factors are also associated with elevated risk of harm. Not all of those experiencing self-harm ideation have a history of mental health challenges and may instead be experiencing adverse circumstances. However, Ofcom OET data suggests that internet users with mental health conditions are more likely to report encountering content promoting self-harm[2082] (7% vs 3% with no conditions). Other evidence also suggests that those with existing mental health concerns may be more likely to encounter suicide or self-harm content.[2083]

20.44 Additionally, the **ethnicity**, **sexual orientation** and **gender** of users may play a role in increasing the risk of harm related to self-harm content.

20.45 Ofcom's OET data suggests that content promoting self-harm is more likely to be experienced by internet users who are of mixed/multiple ethnicities (9%) or Black (6%) compared to those who are white (3%).[2084]

[2077] Instagram, 2021, Bad Experiences and Encounters Framework (BEEF) Survey. [accessed 6 November 2024].

[2078] According to the Samaritans, evidence suggests content presenting self-harm and suicide behaviours (e.g. an online challenge reported to encourage adolescents and young adults to engage in self-harm and eventually kill themselves in a series of 50 challenges), may be contagious to other users who view this content, with younger people aged up to 24 years old being most susceptible. Source: Samaritans, 2022. Towards a suicide-safer internet. [accessed 13 March 2023].

[2079] In 2011-2018, a national confidential inquiry into suicide and homicide by people with mental illness found that 15% of under-25s (aged 10+) had reported using the internet for suicide-related purposes (e.g. visiting pro-suicide websites) during this time, which was significantly higher than for patients aged 25+ (7%).
Source: University of Manchester, 2021. The National Confidential Inquiry into Suicide and Safety in Mental Health. [accessed 11 July 2023].

[2080] Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. [accessed 10 July 2023].

[2081] Samaritans' response to 2023 Ofcom Call for Evidence: Second phase of online safety regulation: Protection of Children

[2082] Ofcom's OET refers to mental health as 'Anxiety, depression or trauma-related conditions, for example.' Experiences of using online services - Ofcom.

[2083] Ofcom research into how people are harmed online included a case study on a male aged 26-30. He had struggled with his mental health over the lockdown period and as a result had sought information on suicide methods via a search engine, until he reached forums that discussed suicide methods. His poor mental health increased the likelihood that he would experience harm from the content. Ofcom, 2022. How people are harmed online: Testing a model from a user perspective. [accessed 11 July 2023].

[2084] Ofcom, 2024. Online Experiences Tracker. [accessed 22 November 2024].

20.46 OET data also suggests that this online harm is experienced significantly more by adult internet users who identify with 'other' sexuality (10%) and bisexual (10%) vs heterosexual (3%). Furthermore, those who identified as non-binary were significantly more likely to report experiencing content promoting self-harm (17%) than males (4%) and females (3%).[2085]

20.47 According to surveys and hospital admissions data, females engage in non-suicidal-self-injury (NSSI) at a higher rate than men. Rates of hospital admission due to self-harm in children and young people (10-24), for instance, are considerably higher for females than males. In 2021/22, 4.6 times as many females than males aged 10-24 were admitted to hospital due to self-harm (82% of all admissions in 2021/22).[2086] [2087]

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles and fake user profiles

20.48 While anonymity has important benefits,[2088] it can also result in users feeling comfortable in sharing or engaging with more harmful or explicit content, thereby increasing the risk of potential illegal content being shared. Analysis of German language posts depicting non-suicidal self-injury (NSSI) on Instagram in 2016 found that around 80% of the posts were posted by anonymous accounts, those displaying no personal information.[2089]

20.49 In other cases, anonymity can embolden users to intentionally encourage others to acts of self-injury, with less risk of being identified and facing consequences. An analysis of self-harm communities on X (formerly Twitter) revealed that some accounts, which claimed to be children, demonstrated the sophisticated operational security of 'internet-savvy adults' These accounts employed advanced techniques to avoid automatic risk-scoring and flagging mechanisms.[2090] This suggests that predators are exploiting self-harm communities for the purposes of exploiting vulnerable children. This is particularly concerning given the existence of networks of accelerationist terrorist groups extorting children online to self-harm.[2091]

[2085] Ofcom, 2024.

[2086] Please note that these data do not include admissions to A&E. Self-harm admission rates also sharply declined, for male and female patients, in 2022/23, but this may be due to a change in NHS England's reporting methodology, which reclassified same day emergency cases, leading to fewer admissions being recorded. Source: Nuffield Trust, 2024. Hospital admissions as a result of self-harm in children and young people. [accessed 01 October 2024]

[2087] Between 2012/13 and 2021/22 the rate for females rose from 508 admissions per 100,000 to 711. For the same period the inverse trend has been observed for males as the rate of admissions decreased by 22% from 193 per 100,000 to 154. Source: Nuffield Trust, 2024.

[2088] Anonymity can have benefits in helping some individuals feel more able to express themselves online, particularly users who may be experiencing thoughts of suicide or self-harm.

[2089] Brown, R. C., Fisher, T., Goldwich, A. D., Keller, F., Young, R., Plener, P. L. 2018. #cutting: Non-suicidal self-injury (NSSI) on Instagram. *Psychological Medicine*, 48, 337-346. [accessed 10 October 2024].

[2090] [DISTRESSING CONTENT WARNING] Network Contagion Research Institute, 2022. Online Communities of Adolescents and Young Adults Celebrating, Glorifying, and Encouraging Self-Harm and Suicide are Growing Rapidly on Twitter. [accessed 10 October 2024]

[2091] Winston, A. (2024). There Are Dark Corners of the Internet. Ten There's 764. *WIRED*. [accessed 10 October 2024]; Argentino, M.-A., Barrett, G., Tyler, M. B., (2024). 764: The Intersection of Terrorism, Violent Extremism, and Child Sexual Exploitation. [accessed 10 October 2024]

20.50 A user's posts on an anonymous user profile can sometimes become more explicit as interest in the profile and associated content grows. One participant in a study by Biddle *et al* (2018) noted "*I created an anonymous Instagram page. At first it was captions and quotes and stuff that I'd find and I thought were quite good… then once I saw how many people were looking at the page I started posting pictures of [self-harm] and getting more and more followers and it became addictive. It eventually got shut down… it became pro-self-harm*".[2092]

20.51 Being able to create a fake user profile in these same circumstances is also considered to increase the risk of content encouraging self-harm being disseminated on a service.

User networking

User groups

20.52 User groups that are dedicated to discussing self-harm topics with other users can be created, particularly on social media services. These can be a source of support, but evidence suggests that they can also contain content which glorifies or normalises self-harm. User groups within social media services are often not moderated in the same way as support forums, where rules about inappropriate content are more likely to exist.[2093]

User communication

Posting content (text, images, videos)

20.53 The ability to post content is an important functionality mentioned in the research and literature on suicide and self-harm. It enables users to communicate and establish contact with others who are experiencing similar thoughts or behaviours, but the evidence shows they it is also being used to negatively influence users' thinking around self-harm.

20.54 A UK-based qualitative study (where participants had either previously used the internet for suicide-related purposes or had been admitted to hospital following serious self-harm) found that among the self-harm patients, most had avoided generating online dialogue and instead preferred to observe others' posts. The study explained that almost all had viewed others' posts on online services about self-harm methods and had used these as a source of information that they could search to gain insight into experiences with these methods, or to decide on details of implementation.[2094]

20.55 Another paper identified that graphic images and videos posted online were, in some cases, found to be emotionally disturbing by people with a history of self-harm and "potentially triggering of self-harm behaviour".[2095] The studies covered various types of posts such as images of wounds and scars, self-harm memes, videos with NSSI (non-suicidal self-injury) content, suicide images from a first-person and third-person perspective, and content

[2092] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.12, *PLoS ONE,* 13 (5). [accessed 10 July 2023].
[2093] Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., Purdy, N., Daine, K. and John, A., 2017. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown, p.16, *PLoS ONE,* 12 (8). [accessed 10 July 2023].
[2094] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018.
[2095] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023. Research Review: Viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms p.17, *Journal of Child Psychology and Psychiatry,* 64 (8). [accessed 10 July 2023].

containing images of self-harm coupled with negative words (such as 'suicide' and 'death').[2096]

Commenting on content

20.56 Potentially illegal pro-self-harm messaging can be found in the comments sections on posted content, evidence from four studies that comments on posted self-harm were used to positively reinforce the sharing of images. This was either by expressing admiration, complementing self-inflicted injuries, or encouraging the original poster to create and post similar images.[2097]

20.57 Analysis of German language posts depicting non-suicidal self-injury (NSSI) on Instagram in 2016 found that while the majority of the comments were either positive (42.54% warning, offering help or empathetic), or discussions (50.1%), that a small but significant number were negative (7.35% abusive or complimentary).[2098] Following this study, 59 individuals who had posted self-generated self-harm images to the Instagram accounts identified in the study discussed above were recruited for qualitative interviews.[2099] While many participants reported being offered help (N=29, 39.2%), only 15.2% believed that those interactions had actually been helpful. By contrast 40.7% (N=24) of the participants reported receiving negative comments including harassment, abuse, and encouragement of suicide. Furthermore, 39% (N=29) reported feeling angry or sad upon receiving negative comments. Participants also reported receiving problematic positive reinforcement of their self-harming, with 5 reported receiving compliments including comments that their wounds were "cool" or "beautiful". A recent analysis of communities sharing self-harm content on X found that comments were used to express admiration for severe self-injury and to provide advice on how to injure themselves more severely.[2100]

20.58 Comments can foster competition to self-harm more frequently and more severely. NSSI posts on Instagram depicting wounds generated around twice as many comments from other users than pictures not depicting wounds. There was also a significant positive association between the severity of the wound depicted and the number of comments.[2101] A review of additional qualitative studies found evidence that competition to increase the severity of self-inflicted injuries is spurred by receipt of negative comments from others.[2102]

Reacting to content and re-posting or forwarding content

20.59 Validation from other users on a service, through means such as 'likes', comments or re-posting, can reinforce or even exacerbate negative thought patterns or behaviours (and potentially encourage the further posting of potentially harmful content). It can also

---

[2096] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023.
[2097] Susi, K., Glover-Ford, F., Stewart, A., Knowles Bevis, R. and Hawton, K., 2023.
[2098] Brown, R. C., Fisher, T., Goldwich, A. D., Keller, F., Young, R., Plener, P. L. 2018. #cutting: Non-suicidal self-injury (NSSI) on Instagram. *Psychological Medicine*, 48, 337-346. [accessed 10 October 2024].
[2099] Brown, R. C., Fischer, T., Goldwich, D. A., & Plener, P. L. (2020). "I just finally wanted to belong somewhere"—Qualitative Analysis of Experiences With Posting Pictures of Self-Injury on Instagram. *Frontiers in Psychiatry*, 11, 274. [accessed 18 November 2024]
[2100] [DISTRESSING CONTENT WARNING] Network Contagion Research Institute, 2022. Online Communities of Adolescents and Young Adults Celebrating, Glorifying, and Encouraging Self-Harm and Suicide are Growing Rapidly on Twitter. [accessed 10 October 2024]
[2101] Brown, R. C., Fisher, T., Goldwich, A. D., Keller, F., Young, R., Plener, P. L. 2018.
[2102] Marchant, A., Hawton, K., Burns, L., Stewart, A., & John, A. (2021). Impact of web-based sharing and viewing of self-harm–related videos and photographs on young people: Systematic review. *Journal of medical Internet research*, 23(3), e18048. [accessed 10 October 2024].

provide users with a sense of community in feeling that they are not alone in their thinking.[2103]

Group messaging

20.60 Group messaging allows users to contact one another and potentially encourage harmful behaviour in a group setting.[2104] Research with adolescents with repeated nonfatal self-harm experiences identified online 'chat groups' as a source of self-harm content, and somewhere other users encouraged people to self-harm.[2105]

Livestreaming

20.61 There have been cases of livestreaming functionalities being used to show users self-harming in real time. This is considered to be of particular concern as those hosting the livestream may be in an extremely vulnerable situation, and other users (viewers) reactions to or engagement with the content could act to further encourage acts of self-harm. Group messaging and commenting functionality, which would enable this interaction, is explored further in the sections below.

Content exploring

Content tagging

20.62 The ability to tag content, such as hashtags, are also a risk factor for disseminating self-harm related content. Variations of suicide and self-harm-related hashtags may be used to avoid content removal. These hashtags can often create spaces where harmful content can proliferate for extended periods without detection by online services. The use of some hashtags to disguise the true nature of suicide and self-harm content may also increase the risk that more users will unintentionally encounter this content.[2106]

20.63 In other cases, variations of hashtags that are likely to be blocked have been used in an attempt to continue to access the content. Despite some effort by services to remove hashtags associated with suicide and self-harm,[2107] it is still possible to retrieve large amounts of potentially harmful suicide and self-harm related content on Instagram. Using hashtags linked to content viewed by Molly Russell prior to her death, the Molly Rose Foundation were able to compile a list including newly created hashtags linked to suicide and self-harm material. With these hashtags they were able to retrieve large quantities of publicly available potentially harmful suicide and self-harm related content on Instagram and TikTok.

[2103] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.12, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[2104] Please note that research often references 'chatrooms' rather than 'group messaging'. Because chatrooms are centred around enabling users to message one another in groups, we have used research on chatrooms to draw conclusions surrounding group messaging.

[2105] Chen R, Wang Y, Liu L, et al. 2021. A qualitative study of how self-harm starts and continues among Chinese adolescents. *BJPsych*. [accessed 01 October 2024].

[2106] Arendt, F., Scherr, S. and Romer, D., 2019. Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults, p.3, *New Media & Society,* 21 (11-12). [accessed 10 July 2023].

[2107] Among other hashtags identified by the researchers as being commonly associated with suicide and self-harm elated content.

User-generated content search filtering

20.64 In some cases, users can apply filters when they search for user-generated content on a U2U service to remove or avoid supportive content (e.g. links to support service) that may support the user who is having suicidal or self-harm thoughts.

20.65 Participants in a UK research study (who had either previously used the internet for suicide-related purposes or had been admitted to hospital following serious self-harm) described how they would 'sift through' user-generated content. The study found that some participants would actively avoid or block out online help (such as pop-ups and support links) once their suicidal thoughts became intense and would filter user-generated data to remove support-giving responses.[2108]

Hyperlinking

20.66 Hyperlinks may contribute to the risks of harm related to suicide and self-harm content. Some studies have shown that hyperlinks can cause a 'rabbit-hole' effect, whereby users engage with links to similar content, leading them to more harmful content which they had not necessarily set out to view.[2109]

20.67 A study on suicide-related internet use found that many young adults in their sample would follow links within and across services. The study found that this behaviour tended to increase as mood lowered, leading to an escalation in browsing and exposure to issues that the participants had not previously considered.[2110]

Saving Content

20.68 The Molly Rose foundation found that, in some cases, potentially harmful suicide and self-harm content has been saved by large numbers of users.[2111] The Coroners inquest into the death of Molly Russell revealed that large quantities of content that she has saved were depression, self-harm or suicide-related.[2112] Saving suicide and self-harm related content may increase frequency of exposure, facilitate rumination, and increase the risk presented by highly depressive content that becomes harmful when viewed in large quantities over time.

[2108] These participants said that by this point, they had decided that they wanted to end their life and were online to research how to action it, looking only for this type of user-generated content. Source: Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.11, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[2109] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, p.8, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[2110] Biddle, L., Derges, J., Goldsmith, C., Donovan, J L. and Gunnell, D., 2018.

[2111] Molly Rose Foundation and The Bright Initiative. 2023. Preventable yet pervasive: The prevalence and characteristics of harmful content, including suicide and self-harm materials, on Instagram, TikTok and Pinterest. [accessed 10 October 2024].

[2112] BBC, 2022. Molly Russell: Social media causes no end of issues, head says. [accessed 10 October 2024]

Recommender systems

Content recommender systems

20.69 Some evidence suggests that content recommender systems[2113] can increase the risk of exposure to self-harm-related content. As recommender systems are understood to maximise user engagement, they can make it more likely that users who engage with harmful content see more of it in the future. In a national survey by Swansea University and Samaritans (where 87% of the sample reported having self-harmed before), over four in five (83%) respondents reported coming across self-harm and suicide content through feeds of recommended content on social media, despite not having searched for it.[2114]

20.70 Recommender systems are capable of curating continuous or limitless feeds of content that can enable episodes of binge-watching behaviour, which is likely to have a negative impact on vulnerable individuals. If a user is primarily engaging with harmful content, they are likely to find more harmful content in their recommender feeds, and therefore have a higher risk of encountering such content during an episode of binge-watching.[2115]

20.71 Therefore, while pieces of content judged in isolation may not be considered illegal, there is significant potential for a risk of harm from the cumulative impact amounting from sustained exposure to suicide and self-harm-related content propagated by recommender algorithms.[2116] In a recent study, the Molly Rose foundation discovered that it was possible to train Instagram reels' recommender system to serve up a sequence of videos such that almost all were considered by them to be potentially harmful. They considered 13% of the videos from this sample to promote and glorify suicide and self-harm.[2117]

20.72 Recommender systems also have the potential to display suicide or self-harm content to those who may not have previously engaged with it. Researchers from the Centre for Countering Digital Hate created four TikTok accounts with female usernames registered as 13-years-old in the USA, the UK, Australia and Canada.[2118] They discovered that these accounts were recommended self-harm, suicide and eating disorder content by TikTok within minutes of scrolling the 'for you' feed. At the earliest, suicide content appeared within the first 2.6 minutes, with eating disorder content recommended within 8 minutes.[2119]

[2113] In the context of online services, a recommender system (or a recommender engine) is a type of information retrieval and ranking system that curates content to a service user. Recommender systems are powered by a set of algorithms which, depending on what they are optimised for, set the decision path for what content is suggested to the user. The goal of a recommender system is to generate recommendations likely to engage the user, although the exact metric/goal will vary by platform.

[2114] Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. p.4. [accessed 10 July 2023]

[2115] University of Sheffield (Dr Ysabel Gerrard), How we're helping social media companies remove harmful content and protect their users. [accessed 10 January 2023].

[2116] The Coroner's Service, 2022. Prevention of Future Deaths. [accessed 28 October 2022].

[2117] Molly Rose Foundation and The Bright Initiative. 2023. Preventable yet pervasive: The prevalence and characteristics of harmful content, including suicide and self-harm materials, on Instagram, TikTok and Pinterest. [accessed 10 October 2024].

[2118] Across all accounts, researchers expressed an interest in body image, mental health and eating disorders by watching and liking relevant videos.

[2119] Centre for Countering Digital Hate, 2022. Deadly By Design: TikTok pushes harmful content promoting eating disorders and self-harm into users' feeds p.19. [accessed 11 July 2023].

20.73 Recommended search terms or search term completions can also exacerbate risk in two ways. The first is by amplifying prior tendencies to view harmful suicide and self-harm related content by reducing friction in the search process. The second is by introducing users to novel search terms that may surface more harmful content that the user would not otherwise have seen. For example, after simulating a proclivity for viewing suicide and self-harm related content, The Molly Rose Foundation discovered that avatar accounts on TikTok were serving up problematic recommended search terms "*people also search for 'quickest way to end it', 'I am going to end it soon', 'I am going to end it'*", and autocompletions for the prompt string 'want to' which included 'end it', 'cut', 'give up' and 'go missing'.[2120]

20.74 The Molly Rose Foundation discovered that publicly available suicide and self-harm related content can receive high levels of engagement. 54% of potentially harmful suicide and self-harm related content that they discovered on TikTok had been viewed over 1 million times. Given the nature of the platform and the scale of the viewer figures, it is likely that this content has been amplified by TikTok's content recommender systems.[2121]

Risk factors: Business models and commercial profiles

20.75 There is some evidence to suggest that advertising-based revenue models may be a risk factor for suicide and self-harm content. In its 2023 Protection of Children Call For Evidence (CFE) response, the Molly Rose Foundation noted that email and push notifications can direct children to suicide and self-harm content. These are sent to users to encourage continued engagement with a service provider to drive up advertising revenue, increasing the risk by encouraging a user to revisit potentially harmful recommended content that the user may have previously engaged with.[2122] Some evidence suggests that there are instances where this revenue model can suggest further suicide and self-harm content to an online user.[2123]

[2120] Molly Rose Foundation and The Bright Initiative. 2023.
[2121] Molly Rose Foundation and The Bright Initiative, 2023. This was a limited study, and as such its findings are not representative of a user experience on the platform. Nonetheless, it does demonstrate that it is possible for a very high volume of potentially harmful content to be served up to a user.
[2122] Molly Rose Foundation response to 2023 Ofcom Call for Evidence.
[2123] One example provided was an email sent to Molly Russell before she took her own life. The Call For Evidence response states that this email contained images of self-harm (some of a graphic nature), suicide (including methods) and depression. Source: Molly Rose Foundation response to 2023 Ofcom Call for Evidence.

21. False communications

Summary analysis for false communication offence: how harm manifests online and risk factors

A person commits the false communications offence if they send a message, with no reasonable excuse to send it, that they know to be false and intend for that message to cause harm.

The risks of harm to individuals are broad. Some individuals might experience distress and anxiety due to false communications shared with the intention to cause harm. Physical harm can also be caused by false communications, for instance from a hoax bomb threat.

The rapid pace of development in generative AI technologies and models presents both a risk and an opportunity. At present, it appears that generative AI technologies and models are likely to enhance the risks of false communications, such as by breaking down barriers to distributing content, and reducing the costs of creating persuasive, false content, particularly through the creation and dissemination of deepfakes.[2124]

Service type risk factors:

Social media services were identified as carrying higher risks of harm for false communications. These services can be used to spread disinformation in foreign influence operations, forms of which could be relevant to how the false communications offence can manifest. This is also true of **messaging services** with encryption, which can be used to spread disinformation due to their closed and encrypted nature.

User base risk factors:

Research suggests that **gender, religious affiliation** and **ethnicity** of users can be risk factors. Disinformation campaigns can often be gendered, with disinformation campaigns targeting women in power more often than men and may also target religious affiliations. Research shows that **diaspora communities** can also be particularly at risk of being victims of disinformation which could include false communications, falling within the parameters of this offence.

Functionalities and recommender systems risk factors:

[2124] Deepfakes are a specific type of media that involves the use of AI algorithms, particularly generative AI models, to modify videos, images, or audio to create realistic synthetic content. This is often done by superimposing the face of a person onto the body of another person in a video or image, as well as voice manipulation with lip syncing. Deepfakes are shared as user generated content on user-to-user services but could also potentially be created using functionalities present on user-to-user services. Deepfake technology is currently used to create content that can be harmful; however, we acknowledge that it may also have positive use cases.

Fake user profiles can be created by perpetrators of foreign influence operations to hide their identity and impersonate authoritative and high-profile sources, through which they can share false information. The ability to post content anonymously, which can be achieved by creating an **anonymous user profile,** can also be exploited in foreign influence operations. We believe that this evidence will also apply to the false communications offence.

Research shows that the ability to **post content** is essential to this offence because it enables disinformation, and potentially false communications, to be disseminated. **Direct messaging** and **encrypted messaging** are other avenues that perpetrators may use to spread false communications. The ability to **edit visual media,** such as creating deepfakes, can also be exploited by perpetrators of the false communications offence.

Introduction

21.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- Content on user-to-user (U2U) services that may amount to the false communications offence listed under 'Relevant offences'; and
- The use of these services for the commission and/or facilitation of this offence (collectively the 'risks of harm').

21.2 We set out the characteristics of U2U services that we consider are likely to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

Relevant offences

21.3 In this chapter, we consider the specific non-priority offence of sending false communications, created by section 179 of the Online Safety Act (the Act). It is an offence for a person to send a message which conveys information that they know to be false, if at time of sending that message they intended the message, or the information in the message, to cause non-trivial psychological or physical harm to a likely audience, and they have no reasonable excuse for sending the message.[2125]

21.4 The offence can also be committed by a person who forwards or shares another person's message or post. An example of false communications would be a hoax bomb threat.[2126]

[2125] Section 180 of the Actsets out a number of exemptions from the false communications offence created. For example, a recognised news publisher cannot commit this offence and therefore content will not be illegal where it has been posted by a recognised news publisher.
[2126] UK Government (Department for Digital, Culture, Media and Sport), 2022. Online safety law to be strengthened to stamp out illegal content. [accessed 18 November 2024].

21.5 For more details on the offence and how services can assess whether content amounts to illegal content, refer to the Illegal Content Judgements Guidance ('ICJG').

How false communication manifests online

21.6 This section is an overview which looks at how the false communications offence manifests online, and how individuals may be at risk of harm.

21.7 The UK Government created the false communications offence as part of an update to the existing offence in the Communications Act which captured knowingly false communications.[2127] It aims to cover "false communications deliberately sent to inflict harm, such as hoax bomb threats, as opposed to misinformation where people are unaware that what they are sending is false, or genuinely believe it to be true".[2128]

21.8 As this is a new offence created by the Act, in preparing this risk assessment we have considered evidence of conduct that appears to be broadly aligned with the conduct intended to fall within the scope of the offence. Using the evidence, we have drawn inferences about the characteristics of services that may be relevant to the risks of harm to individuals of this offence.

Risks of harm to individuals presented by the false communication offence

21.9 While there is limited evidence for how this offence is likely to manifest online, on assessment of this evidence we believe that it will be comparable with the foreign interference offence (see the Foreign interference chapter for more information) and with some of the offences discussed in the chapter on fraud (see Fraud and financial services offences chapter for more information).

21.10 Multimodal disinformation, which is used in the foreign interference offence, combines image and text formats to create false content. These techniques may be exploited by perpetrators of the false communication offence. There are four key methods:

a) De-contextualisation: When real images or videos are paired with false or manipulated text.
b) Reframing: When videos are cropped or decontextualised to make certain aspects or issues more obvious or prominent in pursuit of a specific agenda.
c) Visual doctoring: When images or videos are manipulated to present a different reality to that in their non-edited form – this covers both cheapfakes and deepfakes.
d) Multimodal doctoring: When content is fabricated by pairing manipulated images or videos with false or manipulated content.[2129]

[2127] The false communications offence will replace the offences in section 127(2)(a) and (b) of the Communications Act 2003 and section 1(a)(iii) of the Malicious Communications Act 1988, which will be repealed by section 189 of the Online Safety Act.
[2128] Department for Digital, Culture, Media and Sport, and Home Office, 2022. Online safety law to be strengthened to stamp out illegal content. [accessed 20 September 2023].
[2129] Hameleers, M., Powell, T.E., Van Der Meer, T.G.L.A, and Bos, L., 2020. A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated on Social Media. *Political Communication*, 37(2), pp.281-301. [accessed 20 September 2023]. (US study)

21.11 The evidence we have assessed suggests that the deployment of bots[2130] can be exploited by perpetrators of the false communications offence. Under the direction of a person, they can generate or amplify content which is intended to cause harm to a specific audience.

21.12 Bots are typically employed on social media services to simulate human behaviour. They are often used in foreign influence operations and can be used for like-farming[2131] and click-farming[2132], hashtag hijacking[2133], initiating a repost storm (when a post is instantly reposted by a network of users) and trend-jacking.[2134] [2135] We believe that bots, including those which might employ GenAI technologies, could be used similarly by perpetrators of the false communication offence.

21.13 We expect that false communications are made to elicit a response from someone. This may be to influence them into undertaking a certain activity or to cause psychological harm to an individual, such as fear and anxiety.

Evidence of risks of harm on user-to-user services

21.14 We consider that the risk factors below are liable to increase the risks of harm relating to the false communications offence.

Risk factors: Service types

21.15 Research indicates that the following types of services can be used to facilitate or commit the false communication offence: social media services and messaging services.

Social media services

21.16 Social media services can be used by potential perpetrators to spread disinformation. For example, social media services were used to disseminate false claims related to a COVID-19 vaccine in 2020 that originated as "misleading articles and petitions" on other services. Our evidence shows that disinformation can often originate on more anonymous or closed online spaces before being shared on open social networks, such as some social media services.[2136] An in-depth investigation led by the European Centre for Disease Prevention and Control investigated many misinformation campaigns surrounding the COVID-19 vaccine and found some were being spread with malicious intent – these could cause harm.[2137] This content could then be picked up and amplified by organisations with a

[2130] 'Bots' is an umbrella term that refers to a software application or automated tool that has been programmed by a person to carry out a specific or predefined task without any human intervention.
[2131] 'Like-farming' refers to the use of fake pages on social media sites designed to artificially increase the popularity of a page, so it can be sold to buyers seeing accounts with large followings or for scam and fraud activity.
[2132] 'Click-farming' refers to the practice of manually clicking on online adverts to increase the clickthrough rate value, boost engagement metrics, and inflate impressions. This activity can be carried out by bot accounts, as discussed here, or by large groups of workers.
[2133] 'Hashtag hijacking' refers to the use of a hashtag for a purpose other than it was created – such as tagging a message containing undesirable or harmful content with a popular, but unrelated, hashtag to surface this content to a target audience.
[2134] 'Trend-jacking' refers to when influencers, brands or organisations insert themselves into conversations online that are gaining a lot of attention – for example, by using associated hashtags or trending audios.
[2135] US Department of Homeland Security, 2018. Social Media Bots Overview.
[2136] Facebook, 2021. July 2021 Coordinated Inauthentic Behavior report. [accessed 1 September 2023].
[2137] European Centre For Disease Prevention And Control, 2021. Countering online vaccine misinformation in the EU/EEA. [accessed 20 September 2023].

greater potential reach, thereby increasing the number of users who could encounter it (see Posting content sub-section for more information).

21.17 Bots programmed to spread disinformation are often used on social media services to carry out techniques such as like- and click-farming in foreign influence operations. This could be similar to the way that bots are used by perpetrators of the false communication offence.

Messaging services

21.18 Our evidence also highlights the risk that messaging services with encryption pose in the spread of disinformation, to diaspora communities in particular. This suggests that these services can be used by perpetrators of the false communication offence. The research highlights that their 'encrypted and closed' nature makes fact-checking and content moderation challenging, and consequentially, "these platforms have become a promising new avenue for the spread of disinformation, particularly among diaspora communities"[2138] (see Risk factors: user base sub-section for more information).

Risk factors: user base

User base size

21.19 There is no evidence to indicate that user base size is a specific risk factor for the false communications offence. However, we expect that the number of users on a service could play a similar role as that presented in 'Understanding illegal harms online' in the introduction of the Register of Risk.

User base demographics

21.20 The following section outlines the key evidence on user base demographic factors and the risks of harm, which can include protected characteristics. Services should consider the intersecting influence of demographic factors on risk, which can be contextual, complex and involve multiple factors.

21.21 The data suggest that the user base characteristics of gender, diaspora communities, and religion could lead to increased risks of harm to individuals.

21.22 Gendered disinformation has been defined as information activity that includes creating, sharing and disseminating content which attacks or undermines people based on their gender, or weaponises gendered narratives to promote political, social or economic objectives. Women are often the target of these campaigns.[2139] Although we are still unsure about the conduct that would constitute this offence, perpetrators of the false communications offence may use similar tactics.

21.23 A gendered disinformation campaign under this offence could look like the one which targeted Ukrainian MP Svitlana Zalishchuk after she gave a speech to the United Nations on the effect on the country of Ukraine's war with Russia. This campaign included a screenshot of a falsified tweet claiming that she had promised to run through Kyiv naked if the

---

[2138] Gorksy, J., Riedl,M.J, and Woolley, S., 2021. *The Disinformation Threat to Diaspora Communities in Encrypted Chat Apps*, Tech Stream, Brookings Institute. [accessed 18 November 2024].

[2139] Demos (Judson, E., Atay, A., Krasodomski-Jones, A., Lasko-Skinner, R., and Smith, J.) and the US National Democratic Institute, 2020. *Engendering Hate: The Contours of State-Aligned Gendered Disinformation Online*. [accessed 6 March 2023].

Ukrainian army lost an important battle, and doctored images, purporting to show her doing so.[2140]

21.24 During the 2020 US Presidential election, a hoax about 'disregarded ballots' spread online using encrypted private messaging services, particularly among diaspora communities.[2141] It is reasonable to assume that a false communications offence may include the spread of hoaxes across diaspora groups in a similar way.

21.25 Religious affiliations can sometimes be the target of disinformation, and perhaps false communications fitting within the parameter of this offence. During the COVID-19 pandemic, the BBC found evidence that some disinformation on COVID-19 vaccines was targeted towards individuals with specific religious affiliations, utilising known narratives linked to vaccine hesitancy, with messages falsely claiming that the vaccines contained animal products. Claims targeting Muslims suggested that the vaccines contained pork, while claims targeting Hindus claimed that they contained beef.[2142]

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

21.26 Research has shown that fake user profiles are often created to spread harmful false information as part of foreign influence campaigns. This suggests that they can also be used by perpetrators of the false communications offence. Fake user profiles can be created to hide the identity and impersonate authoritative and high-profile sources which share false information (see Foreign interference chapter for more information).

Anonymous user profiles

21.27 The evidence we have assessed suggests that the ability to post content anonymously can be exploited by perpetrators of foreign influence operations, and we believe that this evidence will also apply to the false communications offence.

21.28 Users involved in disinformation campaigns have exploited anonymity on services by either creating an anonymous user profile, or using services which allow posting content anonymously without an account.[2143]

User communication

Posting content

21.29 The evidence we have assessed suggests that the ability to post content online is key to the false communications offence.

21.30 Think-tank and campaign group First Draft's article 'Trumpet of Amplification' highlights how individuals spreading disinformation have used encrypted and anonymous online

---

[2140] Jankowicz, N., 2017. How disinformation became a new threat to women. [accessed 6 March 2023].
[2141] Gorksy, J., Riedl, M.J, and Woolley, S., 2021. *The Disinformation Threat to Diaspora Communities in Encrypted Chat Apps.* Tech Stream, Brookings Institute. [accessed 9 October 2023].
[2142] Kotecha, S., 2021. *Covid: Fake news 'causing UK South Asians to reject jab,* *BBC News,* 15 January. [accessed 9 June 2023].
[2143] RAND Corporation, (Cohen, R.S., Beauchamp-Mustafaga, N., Cheravitch, J., Demus, A., Harold, S.W., Hornung, J.W., Jun, J., Schwille, M., Treyger, E, and Vest, N.), 2021. Combatting Foreign Disinformation on Social Media. [accessed 20 September 2023].

spaces to create rumours and place fabricated content. This content starts in encrypted spaces before moving to conspiracy communities in closed and semi-closed networks, and then onto mainstream social media services. It is presumably posted on each of these spaces. This content is finally reported on by professional media sources, which might not always have "the training to deep dive into the provenance of the posts, images or videos they find online."[2144]

21.31 In late 2020, a network of accounts on several social media services "posted memes and comments" which spread the false claim that the AstraZeneca COVID-19 vaccine was dangerous because it was derived from a chimpanzee adenovirus.[2145]

21.32 While there are opportunities from the use of generative AI[2146], some evidence suggests that the use of it to create content, which when posted on social media services, messaging services, video-sharing platforms and other services, could be used in false communications offences. This content can take a variety of forms, including audio, video, images and text.

21.33 The evidence outlined in this section does not focus on confirmed examples of the false communications offence; rather, the evidence shows how generative AI technologies could be utilised to create and disseminate content by perpetrators of a possible false communications offence. Not all content created via generative AI will be relevant to the false communications offence, and content that may be relevant to the false communications offence may also be relevant to other harms, such as fraud and foreign interference.

21.34 British politicians have been targeted by deepfake audio and video content, with an audio clip allegedly depicting the then-Leader of the Opposition Sir Keir Starmer swearing at a member of his staff posted to several social media services in October 2023. Fact-checking organisation Full Fact found no evidence that the clip was genuine.[2147] London Mayor Sadiq Khan has also been targeted by deepfake audio, with a scripted, AI-generated replica of his voice used to suggest had made inflammatory remarks ahead of Armistice Day in 2023 and shared widely on social media.

21.35 Additionally, a report by media and research consultancy Fenimore Harper Communications found over 143 deepfake video advertisements impersonating former Prime Minister Rishi Sunak, alongside a series of journalists, on a popular social media service between 8 December 2023 and 8 January 2024, promoting various false investment schemes.[2148]

21.36 An investigation into this network of deepfake videos of politicians and other high-profile figures by the Bureau of Investigative Journalism found that the videos were viewed around one million times in the UK, noting that multiple online reviews have been posted by individuals claiming to have lost money to the false investment scheme. The investigation

[2144] First Draft, (Wardle, C.), 2018. 5 Lessons for Reporting in an Age of Disinformation. [accessed 20 September 2023].
[2145] Facebook, 2021. July 2021 Coordinated Inauthentic Behaviour Report. [accessed 12 June 2023].
[2146] Ofcom, 2024. Ofcom's strategic approach to AI 2024/25, p.3. [accessed 20 November 2024].
[2147] Full Fact, 2023. No evidence that audio clip of Keir Starmer supposedly swearing at staff is genuine. [accessed 16 May 2024].
[2148] Fenimore Harper Communications (Beard, M.), 2024. Over 100 Deep-Faked Rishi Sunak Ads Found on Meta's Advertising Platform. [accessed 16 May 2024].

also found that the advertisements targeting the UK presented the scam as a government initiative and ran across several social media services and video-sharing platforms.[2149]

Direct messaging and encrypted messaging

21.37 Direct messaging represents another means by which false communications can be disseminated. Encrypted messaging can also be used. For instance, research shows that encrypted private messaging services can be used to spread disinformation, particularly among diaspora communities.[2150] It is reasonable to assume that perpetrators of the false communications offence may use services with direct and/or encrypted messaging in a similar way.

Content editing

Editing visual media

21.38 The creation of deepfakes can be exploited by perpetrators of the false communications offence.[2151] In March 2022, a low-quality deepfake of Ukrainian president Volodymyr Zelenskyy talking about Ukrainian armed forces surrendering to Russia was taken down by social media services.[2152] [2153]

Risk factors: Business models and commercial profiles

21.39 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[2149] Bureau of Investigative Journalism (Visser, F. and McIntyre, N.), 2024. Doctored Footage and Hijacked Accounts: Anatomy of a Deepfake Scam Network. [accessed 6 September 2024]
[2150] Gorksy, J., Riedl, M.J, and Woolley, S., 2021. *The Disinformation Threat to Diaspora Communities in Encrypted Chat Apps.* Tech Stream, Brookings Institute. [accessed 18 November 2024].
[2151] Donovan, J, and Paris, B., 2019. Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence. [accessed 20 September 2023].
[2152] BBC News, (Wakefield, J.), 2022. Deepfake presidents used in Russia-Ukraine war, 18 March. [accessed 6 March 2023].
[2153] Gleicher, N., 2022. Tweet on removal of Zelenskyy deepfake, published 16 March 2022. [accessed 6 March 2023].

22. Obscene content showing torture of humans and animals (the s.127(1) offence)

Summary analysis for the offence of obscene content depicting human and animal torture: How harms manifest online, and risk factors

This section considers the evidence of risk factors for the obscene content offence (torture of humans and animals). This can manifest as pre-recorded content (images and video) which can cause significant levels of distress. There are several risk factors which may be associated with the offence, such as the following.

Service type risk factors:

There is some evidence that **social media services**, and to a lesser extent **discussion forums** and **file-sharing and file-storage services** may pose a higher risk of harm connected to this offence, in that they allow users to share content with others, in some cases gaining wide reach, and to engage with the content and other users. Since this content can be publicly available on user-to-user services, it is likely that it can also be discovered through search results on **search services**.

Functionalities and recommender system risk factors:

Any service type which allows the **posting of images or videos**, or the **reposting and forwarding of this content** could be a risk factor for users of online services coming across this content. Similarly, the ability to **share hyperlinks** could mean users risk being exposed to content hosted on other services.

There is some evidence that **content recommendation systems** may play a role in surfacing obscene content online, risking users being harmed despite not actively looking for this type of content.

Business model and commercial profiles risk factors:

It is possible that certain **business models** could be a risk factor for the offence: content creators may be incentivised to post more extreme content, some of which could depict obscene torture of humans or animals, in order to maximise engagement and associated revenues.

Introduction

22.1 This section summarises our assessment of the risks of harm to individuals presented by content on U2U and search services that may amount to the s.127(1) offence listed under 'Relevant offences' below.

22.2 We set out the characteristics of U2U services and, so far as possible, search services, that we consider are liable to increase the risks of harm. 'Harm' means physical or psychological harm. We discuss physical and psychological harm as part of our assessment of the risks of harm, and where possible consider the impact of cumulative harm where content of a particular kind is repeatedly encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

Relevant offence

Torture of humans and animals

22.3 Section 127(1) of the Communications Act 2003 sets out that it is an offence to send, or cause to be sent, online, a message (or other matter) that is grossly offensive or of an indecent, obscene or menacing character where the sender intended, or recognised, at the time of sending, that it may be taken to be grossly offensive, indecent, obscene, or menacing by a reasonable member of the public.

22.4 The s.127(1) offence is a non-priority offence. It has a number of aspects, but we have focused on those aspects of it which are not fully captured by other priority offences (such as Hate, Terrorism and Extreme Pornography), yet give rise to significant public concern. These are depictions of cruelty (in the form of text, images and videos), which are so serious as to be obscene. The offence can be difficult to apply consistently with the right to freedom of expression and providers should therefore have careful regard to our Illegal Content Judgements Guidance.

22.5 This offence is important because for pre-recorded content some illegal animal cruelty and human torture content may not necessarily amount to a priority offence, such as the animal cruelty offence, under the Act due to the specific wording of the offences themselves. However, we recognise that the content may still pose a significant risk of harm to users of user-to-user and search services.

Content harmful to children

22.6 Content which depicts real or realistic serious violence against, or injury of, an animal (real or fictional) is considered priority content that could be harmful to children. The relevant offence, evidence and risks on services is covered in the draft Children's Register of Risk in our second phase of regulation.

How the s.127(1) offence manifests online

22.7 There is evidence that content (text, photos and videos) describing or depicting real – or what appears to be real – torture, serious injury or the deliberate killing of a human or animal does appear on user-to-user services, and is likely searchable through search services, leading to a risk of harm to individuals.

22.8 In this section we are focusing primarily on pre-recorded obscene animal torture content and any obscene human torture content. As explained in the Illegal Content Judgements Guidance and Register of Risks chapters titled 'Animal cruelty', we think that livestreamed content depicting cruelty to animals would be likely to constitute the animal cruelty offence, in that it is a conspiracy between streamers and viewers to encourage and facilitate these acts. We also refer to evidence of pre-recorded content depicting cruelty to

animals in the preceding Register of Risks section on the animal cruelty offence, because its publication may amount to use of the platform for committing the offence and may facilitate further offences.

22.9 However, the most serious pre-recorded content depicting animal torture is also very likely to constitute an s.127(1) offence and pose a risk of harm to users of online services who see it. As regards torture or deliberate serious injury of animals, this is particularly important for search services, which under the Act are not required to consider the risks of their service being used to facilitate or commit the priority animal cruelty offence.

22.10 As regards pre-recorded torture or serious injury of humans, the s.127(1) offence is important for both U2U and search services, because such content may not be caught by the priority offences. It is probable that livestreamed content would amount to a priority offence (threatening behaviour likely to cause fear or alarm)[2154] but this is less clear for pre-recorded content.

Risks of harm to users

22.11 Ofcom's Online Experiences Tracker (OET) suggests that 78% of people in the UK are highly concerned about animal cruelty content.[2155] 10% of people we surveyed had seen animal cruelty content in a four-week period.[2156] Of those who said this was their most recent potentially harmful experience, over three-quarters (77%) said 'It really bothered me/I found it extremely offensive' and a further 22% said 'It slightly bothered me/It slightly offended me'.[2157] This is likely in part out of concern for the animals themselves, but also demonstrates the distress that can be caused by viewing such content.

22.12 The OET also found that, compared to attitudes about – and experiences of – animal cruelty content, a slightly smaller number of people in the UK (74%) are highly concerned about content depicting or encouraging violence or injury,[2158] and a similar number (11%) had seen this type of content in a four-week period.[2159] However, the impact of seeing this type of content was quite different to animal cruelty content: a much smaller proportion of respondents (44%) said 'It really bothered me/I found it extremely offensive', and almost one half of people (50%) said 'It slightly bothered me/it slightly offended me'.[2160]

22.13 In addition, those who said they had seen content depicting cruelty to animals most recently were a more likely to report or flag the content they had seen than those who said they had seen content depicting or encouraging violence or injury most recently (28% compared to 23%). Those who said they had seen violence or injury content were more likely to say they didn't take any action in response to seeing this content compared to those who said they had seen content depicting cruelty to animals (49% compared to 41%).

[2154] Section 38 of the Criminal Justice and Licensing (Scotland) Act 2010.
[2155] Ofcom, Online Experiences Tracker, 2024. [accessed 18 November 2024]. Fieldwork was carried out in a four-week period in May and June 2024. Note that in our November 2023 and August 2024 Consultations we referred to previous iteration of this research (Waves 4 and 5).
[2156] Ofcom, Online Experiences Tracker, 2024.
[2157] Ofcom, Online Experiences Tracker, 2024.
[2158] Ofcom, Online Experiences Tracker, 2024.
[2159] Ofcom, Online Experiences Tracker, 2024.
[2160] Ofcom, Online Experiences Tracker, 2024. For this question the number of respondents was much smaller than others (62), which may mean this is not fully representative.

22.14 In other words, the OET survey might be interpreted to suggest that users' experiences of violence and injury content were potentially less intense or emotive than those of content depicting cruelty to animals. However, this may be because users were not as frequently exposed to the very extreme violence or injury content that would be considered obscene content under the s.127(1) offence.

22.15 Users may also have a higher tolerance for violent content where it involves humans compared to animals; this is potentially evidenced by some recent research into engagement with content depicting cruelty to animals on social media services. Researchers noted that participants were more likely to engage positively with violent human-human content compared to violent human-animal content (albeit not necessarily any content that would constitute an offence).[2161]

22.16 There is evidence that the impacts of viewing violent content could lead to significant levels of distress, even where it is not extremely obscene. In one study, researchers analysed users' discussions of their reactions to seeing violent content (including torture of humans and animals). They found that alongside emotions such as acute and prolonged shock, fear, sadness and discomfort, users also occasionally reported symptoms of physical illness. Some users sought support, but others displayed indifference or even encouraged people to watch it for the purpose of "enhancing – or displaying – psychological resilience".[2162]

22.17 While any illegal content inherently has the potential to cause harm to viewers, there may also be a cumulative impact from repeated exposure to illegal content, such as increased levels (or prolonged periods) of distress.

Evidence of risk factors on user-to-user services

Risk factors: service type

22.18 There is not a great deal of systematic research into the existence and prevalence of obscene torture content towards animals or humans on user-to-user services. However, a broad range of types of user-to-user services can be used to share this content, particularly social media, messaging and discussion forums or chat rooms.

Social media services

22.19 Violent content, such as text describing or videos and images depicting death and serious injury of animals or humans can be found on major social media services, including in publicly available feeds.

[2161] McGuirk, L.Ryan and Alleyne, E. 2024. "Liking," "Commenting," and "Reposting": Psychological factors associated to online animal abuse, *Society & Animals*, pp.14 and 17. Note this study is in pre-print (not yet fully published), and used a non-representative sample through a survey of self-reported attitudes. It should therefore be interpreted only as a potential indication of the variance in attitudes towards human and animal abuse content.

[2162] Stubbs, J. E., Nicklin, L. L., Wilsdon, L., and Lloyd, J. 2024. Investigating the experience of viewing extreme real-world violence online: Naturalistic evidence from an online discussion forum. *New Media and Society*, 26, 3876-3894. [accessed 19 November 2024].

22.20 The Social Media Animal Cruelty Coalition has documented pre-recorded content depicting animal cruelty on social media and video-sharing services,[2163] which we have explored in more detail in the preceding section on the animal cruelty offence. There we suggest it may explicitly encourage, assist or facilitate acts of animal cruelty, and therefore could constitute the animal cruelty offence; we also suggest that this type of content could, by normalising this behaviour, inherently encourage further similar acts even without explicitly doing so. However, where it is obscene torture content, it may also be illegal because of the s.127(1) offence.

22.21 Human torture content has been noted as present in user communities on a social media service[2164] and appearing in recommended videos and on meme pages.[2165]

22.22 The researchers in the study into user reactions to seeing violent content, mentioned above, gathered their data from social media communities in which users discuss their experiences (including seeing content showing torture of humans and animals).[2166] These discussion or descriptions of the content in themselves are unlikely to constitute an offence. However, they may increase the risk that a user is exposed to the content through shared links or through them searching for content after being involved in conversations about it.

Messaging services

22.23 Messaging services may be being used by extremist groups to share explicitly gory and violent videos: research published by Human Digital (in collaboration with the Institute for Strategic Dialogue) identified links to videos were being shared on extremist messaging channels.[2167]

User-to-user pornography services

22.24 user-to-user pornography services may host content showing obscene acts of sexual sadism, including rape or sexual violence. We discuss elsewhere the risk of these services (and messaging services) being used to commit or facilitate the extreme pornography offence.

Discussion forums or chat rooms

22.25 Given the existence of websites specifically for the sharing of so-called 'gore' and 'snuff' image and videos,[2168] it is reasonable to assume that discussion forums or chat rooms are a risk factor for users interested in discussing and sharing this content. Some of this content

---

[2163] Social Media Animal Cruelty Coalition, 2021. Making money from misery. [accessed 19 November 2024]. As noted in the previous section, SMACC's publications are referenced here as an indicator that potential harmful content (including that which could constitute the obscene torture content offence) exists online, not as an indicator of the scale of the issue.
[2164] Dahl, K, 2018. Exploitation of the internet? The morality of watching death online, The Guardian 12 October. [accessed 21 June 2024]; Stephen, B. 2019. Reddit bans r/watchpeopledie in the wake of the New Zealand mosque massacres, The Verge, 15 March. [accessed 21 June 2024].
[2165] Lorenz, T., 2023. Instagram users are being served gory videos of killing and torture, The Washington Post, 26 February. [accessed 21 June 2024].
[2166] Stubbs, J. E., Nicklin, L. L., Wilsdon, L., and Lloyd, J. 2024. Investigating the experience of viewing extreme real-world violence online: Naturalistic evidence from an online discussion forum. *New Media and Society*, 26, 3876-3894. [accessed 19 November 2024].
[2167] Institute for Strategic Dialogue, 2023. Gore and violent extremism: how extremist groups exploit 'gore' sites to view and share terrorist material. [accessed 21 June 2024].
[2168] The Human Digital study noted above analysed 10 of these sites, and include screenshots (images have been removed from the screenshots, but video titles are uncensored and may be distressing).

may be shocking or distressing to some individuals without necessarily being s.127(1) content.

File-storage and file-sharing services

22.26 We have explored elsewhere the use of file-storage and file-sharing services being a risk factor for terrorist material and CSAM, for example. It is therefore reasonable to presume that they are also used to share specifically s.127(1) content.

22.27 The Human Digital study mentioned above characterised 'gore sites' with graphic and violent videos as "proxy file sharing services for extremist and terrorist video content.

Risk factors: user base

User base size

22.28 There is no evidence to indicate that user base size is a specific risk factor for this offence. However, we expect the number of users on a service could be relevant, as described in the 'Introduction to the causes and impacts of online harm'.

User base demographics

22.29 There is some limited evidence to suggest that personal characteristics of users, such as gender, could lead to an increased risk of harm to individuals.

22.30 Some research has suggested gender differences in the experience of viewing unsolicited, real-world, explicit sexual or violent content on online social media services: those identifying as men were finding it less disturbing and more amusing or exciting than those identifying as women.[2169]

22.31 A study into reactions to viewing violent content indicated that there was frequently a gendered component to comments when users were encouraging others to watch this type of content or stopped expressing distress (implying that to be able to watch such videos was a masculine trait).[2170]

Services which allow child users

22.32 Research on the impacts of witnessing content that may be considered obscene torture content has often focused on children.

22.33 The research around exposure to animal cruelty typically comments on it being a risk factor for the child having behavioural problems in future,[2171] or even themselves perpetrating

---

[2169] Nicklin, L. L., Swain, E. and Lloyd, J. 2020. Reactions to unsolicited violent, and sexual, explicit media content shared over social media: gender differences and links with prior exposure, *International Journal of Environmental Research and Public* Health 17. [accessed 19 November 2024]. Note this was a relatively small-scale, non-representative survey (225 survey participants, of whom three-quarters were women), which did not present many options for respondents to choose from when considering their reaction to content. However, it does suggest gender *may* be a factor. The same study also considered the impact of prior exposure to this type of content.

[2170] Stubbs, J. E., Nicklin, L. L., Wilsdon, L., and Lloyd, J. 2024. Investigating the experience of viewing extreme real-world violence online: Naturalistic evidence from an online discussion forum. *New Media and Society*, 26, 3876-3894. [accessed 19 November 2024].

[2171] McDonald et al 2017. The role of callous/unemotional traits in mediating the association between animal abuse exposure and behaviour problems among children exposed to intimate partner violence, *Child Abuse & Neglect* 72, 421-432. [accessed 19 November 2024].

acts of animal cruelty[2172] or other crimes.[2173] This research does not specifically refer to children witnessing animal cruelty online, and in some cases the conclusions are drawn from a small or non-representative sample. Nevertheless, the evidence does suggest that there is potential for harm. Also, given there is evidence that children are seeing this type of content online,[2174] it is a reasonable assumption that this is one mechanism by which they are exposed to it and experience harm. They may also be disproportionately affected by it (particularly psychologically), especially where it is extreme cruelty.

22.34 It is possible for people online, including children, to access murder and torture content on the dark web;[2175] some of these individuals may have moved into dark web spaces after having seen similar content on mainstream areas of the internet and are searching for other, extreme content. Although we have not considered specific evidence about the dark web for the purpose of this section, we acknowledge that it will likely play a role in the online dissemination and exposure to obscene torture content.

Risk factors: functionalities and recommender systems

User networking

User connections

22.35 Services allowing perpetrators to connect and share content with other users is likely a risk factor – as noted above, social media services can host this type of content, including in publicly-available spaces, or it is a way for extremist groups to connect with other users and share illegal content.

Group messaging

22.36 The existence of animal cruelty content being discussed and shared or linked within group messages suggests this is likely to be a risk factor for s.127(1) content.[2176]

User communication

Posting images or videos

22.37 The evidence for a range of animal cruelty content, some of which may be obscene torture content, has been explored in the Register of Risks section on the animal cruelty offence. This includes pre-recorded content which has been identified by the Social Media Animal Cruelty Coalition showing gratuitous and prolonged torture.[2177]

[2172] Thompson, K. L. and Gullone, E. 2006. An investigation into the association between the witnessing of animal abuse and adolescents' behavior toward animals, *Society and Animals* 14, 221-244. [accessed 19 November 2024]. Note that this study is very old, but more recent research suggests similar trends and analyses: for example, Wauthier, L. M. and Williams, J. M., 2022, Understanding and conceptualizing childhood animal harm: a meta-narrative systematic review, *Anthrozoös* 35, 165-202. [accessed 19 November 2024]. This literature review of studies from 2010 until 2020 covers the risk factors for the child as a perpetrator, which includes witnessing animal cruelty itself.

[2173] Johnson, S. A. 2018. Animal cruelty, pet abuse & violence: the missed dangerous connection, *Forensic Research & Criminology International Journal.* [accessed 19 November 2024].

[2174] RSPCA, 2018. *The RSPCA's Generation Kind*, p.2, and footnote 2 [accessed 9 May 2024].

[2175] For instance, it was widely reported that one of the killers of Brianna Ghey, prior to her crime, watched 'real' murders and torture on the dark web: Gawne, E. and PA Media, 2024. Brianna Ghey inquest to look into killer's school transfer, BBC News, 11 April; Cobham, T. 2024 Kill lists, Sweeney Todd and the dark web: how torture-obsessed teenagers plotted Brianna Ghey's murder, The Independent, 2 February. [both accessed 21 June 2024].

[2176] BBC, 2023. Monkey Haters (documentary) [accessed 27 March 2024].

[2177] Social Media Animal Cruelty Coalition, 2021. Making money from misery; Social Media Animal Cruelty Coalition, 2022b. Teasing as Torture, [both accessed 19 November 2024].

22.38 The Human Digital study into extremist videos took them to 'gore' sites which hosted obscene videos. The study notes that of the sites in the sample, none had measures in place to prevent users from seeing the content.[2178]

Reposting or forwarding of content

22.39 The same Human Digital study noted that the sites they viewed allowed users to download videos, allowing for redistribution on other sites or services.

Commenting on content

22.40 We are not aware of any recent evidence for comments and replies on posted content being a particular risk factor for harm from obscene torture content. Where the content being commented on is itself obscene torture content, the comments may reinforce the harm to individuals through further discussion or by encouraging, or directing users to, other content. In one 2017 study on a specific 'shock/gore site' (now defunct), a notable number of users in the comments sections displayed positive emotions towards the content and other users, including humour and a sense of community.[2179] This is likely due, in part, to these users predominantly being like-minded individuals having actively sought out this service and its content.

Content exploring factors

Searching for user-generated content

22.41 Since human and animal torture content may be available on services with user-generated content, including social media services, it stands to reason that any search functionality within this site could be risk factor for seeing this type of content.

Hyperlinking

22.42 The Human Digital study referred to above noted that links to the video-sharing service hosting potentially obscene torture content were being shared within a messaging service, in additional to the content from the service being freely downloading. As with other illegal content which may be shared through hyperlinking to other services or 'content stores' (file-sharing services), such as terror content or CSAM, a hyperlinking feature is likely a risk factor for the sharing of obscene human and animal torture content.

22.43 We have also noted in the 'Animal cruelty' Register of Risks chapter that hyperlinks to private messaging services or file-sharing services may be present in public content shared on social media services.

Recommender systems

Content recommender systems

22.44 Evidence that recommendation systems have played a role in the surfacing of animal cruelty content online – thereby potentially contributing to an online environment in which the unnecessary suffering of animals is encouraged – has been explored in the animal

[2178] Institute for Strategic Dialogue, 2023. Gore and violent extremism: how extremist groups exploit 'gore' sites to view and share terrorist material [accessed 21 June 2024].
[2179] Alvarez, M. 2017. Online spectatorship of death and dying: pleasure, purpose and community in BestGore.com, *Participations: Journal of Audience and Reception Studies*. [accessed 19 November 2024].

cruelty offence section. Some of this animal cruelty content may include obscene animal torture content that would constitute the s.127(1) offence.

22.45 There is some, albeit limited, evidence from media reporting that recommender systems are also a contributing factor for content showing violence and torture of humans: users have reportedly been recommended videos in their main social media feeds, within minutes or even seconds of opening the app.[2180]

Network recommender systems

22.46 While we are not aware of any evidence for network recommender systems being a specific risk factor for the sharing of or exposure to obscene animal or human torture, it is possible that they facilitate user connections between people interested in this type of content, thereby increasing the likelihood of the content being shared to a wider group of users.

Risk factors: Business models and commercial profiles

22.47 As noted in the animal cruelty offence section, there is evidence that animal cruelty videos are being monetised, some of which may be obscene torture content. A service and its content creators may be incentivised to increase user engagement with content in order to increase the revenue earned, by posting ever more extreme content, some of which could constitute the obscene content offence being monetised.

[2180] Lorenz, T., 2023. Instagram users are being served gory videos of killing and torture, The Washington Post, 26 February. [accessed 21 June 2024].

23. Threatening communications

Warning: this chapter contains content that may be upsetting or distressing.

Introduction and relevant offence

23.1 Section 181 of the Online Safety Act creates the 'threatening communications offence' as a new offence to consider in the Register of Risks. In the absence of specific evidence about this new offence, we consider that the risk factors for the threatening communications offence will be the same as those outlined in the harassment, stalking, threats and abuse chapter. We consider them very likely to manifest in the same way.

23.2 A person commits the threatening communications offence if they send a message that conveys a threat of death or serious harm and, at the time of sending it, the sender intended the individual encountering the message to fear that the threat would be carried out, or was reckless as to whether the individual encountering the message would fear that the threat would be carried out. The Act defines 'serious harm' for these purposes to mean:

- serious injury amounting to grievous bodily harm;
- rape;
- assault by penetration; or
- serious financial loss.

23.3 For more details on the offences and how services can assess whether content amounts to illegal content, please refer to the Illegal Content Judgements Guidance or ICJG.

How threatening communications manifest online

23.4 There is currently limited research and evidence available on the threatening communications offence specifically. However, this offence is similar to certain offences listed under the Harassment, stalking, threats and abuse chapter. We have therefore assumed that it will manifest online in the same or similar ways, although we will keep this under review as the evidence base develops.

23.5 We consider that the threatening communications offence can occur on any service that enables users to send a message or communication. This includes social media services, private messaging services, online gaming services, discussion forums and chat rooms, and online dating services.

23.6 To put the risks of harm from this offence into context, a study from the US found a doubling of physical threats received online between 2014 and 2020: from 7% to 14% of US

adults having experienced it.[2181] [2182] Amnesty found that of the one in five women in the UK who had experienced abuse and harassment online, 27% had been threatened with physical or sexual assault.[2183] Evidence also suggests that the prevalence of threatening communications is high among certain groups, such as women in public roles.[2184]

Risks of harm to individuals presented by threatening communications offences online

23.7 Examples of threatening communications online are likely to be similar to content amounting to abuse and harassment, and may include messages, posts, live chat harassment or 'flaming', a form of online verbal abuse.

23.8 The risks of harm to individuals presented by threatening communications offences online are likely to be very similar to those outlined in the harassment, stalking and threats chapter. In particular, this includes the risks of harm arising from aggravated forms of harassment and stalking, and violent threats (which include threats to kill, rape threats and threats of violence) offences. The Register of Risks also refers to this offence in the Controlling or coercive behaviour and Sexual exploitation of adults chapters.

[2181] The 2020 figure comes from a panel of 10,093 US adults. Source: Pew Research. 2020. The state of online harassment. [accessed 28 September 2023].
[2182] Where possible, UK data has been used throughout this chapter. However, when this is limited, evidence for comparable cultures has been used, namely the US, Australia and Canada. Where evidence is not UK-based, this will be clearly stated.
[2183] From a sample of 500 women aged 18 – 55 years old. Source: Amnesty International UK. Online abuse of women widespread in UK. [accessed 28 September 2023].
[2184] Amnesty International, 2018. Toxic Twitter – Women's Experiences of violence and abuse on Twitter', Chapter 3 in Online Violence against Women. [accessed 28 September 2023]; UNESCO (Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S.), 2020. Online violence Against Women Journalists. [accessed 28 September 2023].

24. Search services

Warning: this section contains content that may be upsetting or distressing.

Introduction

24.1 Search services are the starting point of many users' online journeys and play a crucial role in making content accessible. Although they do not host content or enable interaction between users, search services can still cause harm by providing a means for users to locate and access illegal and harmful content.

24.2 Search services are, for the most part, designed to optimise the search experience of their users and help them find the content they are searching for. Many search services do so by indexing webpages from across the 'clear web'.[2185] [2186]

24.3 The risk of encountering illegal content is caused by the fact that any content which has been indexed can be presented in search results and encountered by users unless mitigations are in place that specifically prevent illegal content from being returned in search results.

24.4 We have produced a separate chapter on search services because of the different risk assessment duties that the Online Safety Act (the Act) places on Ofcom and providers in the case of search services, and the fundamental differences between the typical functionalities of search services compared to user-to-user (U2U) services.

Service types

24.5 We refer to search service types that we expect to be recognisable to both users and businesses, to illustrate how harms can manifest online and how the characteristics of a service can affect the risk of harm.

24.6 This chapter sets out the requirements the Act places on Ofcom for the purposes of conducting our assessment of the risks of harm on search services. In the following sections, we have also set out some of the relevant Act definitions in a clear and accessible manner. Where we have described search services, this should not be taken to be a definitive view of the services (or parts of services) that may be in scope of the Act.[2187] It is for services to assess themselves and seek their own independent advice to enable them to understand and comply with the Act. For more, please refer to the Overview of Regulated Services chapter (Volume 1, Chapter 3).

[2185] Indexing is the process of collecting, parsing, and storing of data to facilitate fast and accurate information retrieval.
[2186] The 'clear web' refers to the publicly accessible webpages that can be indexed by search engines.
[2187] A search service is defined in section 3 of the Online Safety Act as an *"internet service that is, or includes, a search engine*". A search engine "*includes a service or functionality which enables a person to search some websites or databases (as well as a service or functionality which enables a person to search (in principle) all websites or databases)"* but "*does not include a service which enables a person to search just one website or database"* (section 229 of the Online Safety Act).

Search services

24.7 Services that allow users to search more than one website or database may be a 'search service'.[2188] Searching can be done by any means, including the input of text, images, videos or speech. Search services are 'regulated' if they fulfil certain requirements, including having a link to the United Kingdom.[2189] A provider of a search service is the entity that has control over the operations of the search engine, which includes operations that enable a user to input a search request and generate responses to those requests in the form of search results.[2190]

24.8 Ofcom has identified several types of search services based on the definitions in the Act and how search services operate.

- General search services: General search services enable users to search the contents of the web by inputting search requests on any topic and returning results. There are two types of general search service:
  - **General search services which rely solely on their own indexing**: These work by using crawlers (also called bots) to find content across the web ('crawling'); building an index of URLs by validating and storing the content found in a database ('indexing'); and using algorithms – for example, Google's PageRank – to rank the content based on relevance to the search query ('ranking'). Search services use many ranking signals, the exact composition of which are proprietary and not necessarily publicly known.[2191] There are a small number of large general search services that do their own crawling, indexing and ranking. Providers of these services may also syndicate some or all of these processes and provide search results to downstream general search services. There are also smaller general search services which do their own indexing.
  - **Downstream general search services:** As a type of general search service, downstream general search services provide access to content from across the web. They do so by obtaining search results from providers of those general search services that conduct their own indexing, and may supplement these syndicated results with additional information and features. The control that a downstream entity has over how search results are displayed on its search service may vary depending on the contractual arrangement with the upstream entity from which it syndicates search results.[2192] Downstream general services often distinguish themselves from upstream general search services by offering a social purpose (e.g. Ecosia), additional privacy (e.g. DuckDuckGo), or differentiated search features.

[2188] Sections 3 and 229 of the Online Safety Act.
[2189] Refer to section 4 of the Online Safety Act and Schedule 1 to the Online Safety Act.
[2190] Section 226(4), (5) and (13) of the Online Safety Act. Section 226 clarifies that there can only be one entity that is the provider of a search service. Please note, as set out in [Approach to Codes chapter] it is for the entities involved in the provision of a search service to seek their own advice as to whether they are the 'provider' of that service.
[2191] The search engine index takes the output from the crawler and creates relevant data structures to support later searching within the search engine. The index can comprise document content, images, and metadata. An index will have many repeated refinement algorithms applied to increase its accuracy and relevance.
[2192] In its advertising market study, the Competition and Markets Authority (CMA) said none of the contracts it had looked at allowed the downstream general search service to re-rank the search results they received from Google or Bing. Source: Competition and Markets Authority (CMA), 2020. Online platforms and digital advertising: Market study final report. [accessed 22 September 2023]. We discuss our position on who the 'provider' of a downstream general search service is in the [Approach to Codes chapter].

- **Vertical search services:** Also known as 'speciality search engines', enable users to search for specific topics, or products or services offered by third party operators with which the provider of the vertical search service has a relevant arrangement. They operate differently from general search services. Rather than crawling the web and indexing webpages, they present users with search results only from selected websites or databases with which they have a contract. An API[2193] or equivalent technical means is used to return the relevant content to users. Common vertical search services include price comparison sites.

24.9 We have also identified the following two ways in which content generated by generative AI (GenAI) models or a GenAI service may fall within scope of the Act's search duties.[2194] First, a standalone GenAI service could constitute a search service in its own right, where it generates output including search results from more than one website or database.[2195] Second, a conventional search service could use a GenAI model to augment its search engine and facilitate the delivery of its search results.[2196] In such cases, the outputs of the GenAI model may constitute search results and be within scope of the Act.

Scope of Ofcom's assessment of risk of harm from illegal content

24.10 This chapter summarises our assessment of the risk of harm to individuals presented by search content on a regulated search service that amounts to illegal content (the 'risk of harm').[2197]

24.11 We set out the characteristics of search services that we consider are liable to increase the risk of harm. Harm means physical or psychological harm[2198] that can occur to an individual as a result of:

- a user directly encountering illegal content in or via the search results[2199] of a search service; or
- harm that can occur to third-party individuals who have not directly encountered illegal content via search results but who may be harmed by the words or actions of those who have.

[2193] Application Programming Interface (API) is a way for two or more computer programs to communicate with each other.
[2194] Please note that this is not an exhaustive list, and service providers should obtain their own legal advice about the ways in which such GenAI content may fall within scope of the OSA's search duties.
[2195] For instance, a GenAI service could draw on more than one website or database by providing real-time information from plug-ins.
[2196] For example, a search service could integrate a GenAI that provides a conversational summary of the results produced by the service's existing search engine.
[2197] Section 98 of the Online Safety Act.
[2198] Section 234 of the Online Safety Act.
[2199] Section 57 of the Online Safety Act defines both 'search content' and 'search results'. 'Search results' includes content presented to a user by operation of the search engine, and 'search content' is content encountered in or via search results. This definition captures content encountered as a result of interacting with search results, for example, by clicking on them and does not include content encountered through subsequent interactions with an internet service other than the search service. Paid-for advertisements, content on the website of a recognised news publisher and other journalistic content is excluded from the definition of 'search content'.

24.12 Content is illegal if it amounts to a relevant offence. This includes both priority offences[2200] and other relevant offences,[2201] including communications offences.[2202]

24.13 'Search content' can consist of words, images, videos, speech, or sound. All these forms of content can constitute illegal content if they amount to a relevant offence. Content may be illegal if using, possessing, viewing or accessing, publishing or disseminating it amounts to a relevant offence.

24.14 For more details on the offences and how services can assess whether content amounts to illegal content, please refer to the Illegal Content Judgements Guidance or ICJG.

How harm manifests on search services

Risk of harm to individuals presented by illegal content on search services

24.15 The role of search services in reducing barriers to accessing information has provided significant benefits to individuals and society. The assessment below does not attempt to weigh up the positives and negatives of these services and the companies that run them. It is only concerned with identifying and assessing the risk of harm. In some cases, such risk of harm is a consequence of the same characteristics that provides benefits in the vast majority of cases.

24.16 Although the mechanisms by which illegal content can be encountered may be different on search services compared to U2U services, we consider that the harm occurring as a result of illegal content being accessed is comparable as if it was encountered on a U2U service. For instance, the existence and distribution of child sexual abuse material (CSAM) causes irreparable harm regardless of where and how perpetrators have been able to access it online. To avoid repetition, we recommend readers refer to the 'Risk of harm' sections in the Part 1: User-to-user services chapters for the corresponding kind of illegal harm, to understand what impacts content that amounts to an offence can have.[2203]

Risks of GenAI in Search

24.17 The rapid integration of GenAI into search services creates a potential for new risks, or new ways for risks of harm to manifest on search services. As noted, there are two ways GenAI can be used to facilitate the delivery of search results to users; search services using generative AI to enhance search results (for example, as an additional feature on the search

[2200] Offences listed under Schedules 5 ('terrorism offences'), 6 ('child sexual exploitation and abuse offences), and 7 (priority offences) to the Act.
[2201] Other offences are defined in section 59(5) of the Online Safety Act and includes all offences under UK law that are not priority offences, where (a) the victim or intended victim of the offence is an individual (or individuals); (b) the offence is created as a result of the Online Safety Act, another Act, an order of Council or other relevant instruments; (c) the offence does *not* concern the infringement of intellectual property rights, the safety or quality of goods, or the performance of a service by a person not qualified to perform it; and (d) the offence is *not* an offence under the Consumer Protection from Unfair Trading Regulations 2008.
[2202] Communications offences are listed in Part 10 of the Act and include false communications offence, threatening communications offence, offences of sending or showing flashing images electronically, and offence of sending etc photograph or film of genitals.
[2203] For example, for a detailed discussion of the harm caused by CSAM, please refer to the CSAM chapter; for a discussion of the harm caused by fraud refer to the Fraud and financial services offences chapter.

results page), and GenAI chatbots that 'search' the internet to provide responses to prompts. We consider both to have the potential to present illegal content to users.

24.18 If the underlying database(s) from which search results are derived contains illegal content, there is a risk it will be presented to users via GenAI tools or services if these are not designed with effective safeguards to prevent this happening (for example, certain prompts triggering a warning message rather than being answered). A further risk is posed by the potential for a GenAI search tool or service to create new illegal content to a user, created for the first time in response to a user prompt. Notably, although tools will usually have guardrails built-in to prevent content being generated in response to certain requests by default, it has been shown these can be reliably bypassed with various methods.

24.19 Although published research doesn't systematically cover every type of illegal harm covered by the Register of Risks, there is evidence that a variety of potentially illegal content can be, or has been, accessed via GenAI on search services. Research has shown search services integrated with GenAI chatbots could be used to facilitate fraud whereby a perpetrator could covertly collect personal information including the user's name, email, and credit card information.[2204] There is also evidence illustrating how such services could be used to share malicious links and steer search results towards manipulated content.[2205]

24.20 We consider the risk of harm resulting from access or exposure to illegal content via GenAI tools or GenAI search services is equivalent to encountering the content on or via a search results page.

Evidence of risk factors on search service

24.21 Evidence reviewed for the Register was concerned with one key area: whether the characteristics[2206] – including the service type, functionalities, business models and user bases – of search services appear to play any role in the risk of harm. This would lead any such characteristic to be considered a risk factor for search services, and likely to increase the risk of harm to individuals.

24.22 All the published evidence that is referenced here is concerned with general search services (including downstream general search services), due to the limited evidence on other types of search services such as vertical search services, as explained in the following paragraph.

Risk factors: Service types

24.23 The ability of users to enter search queries relating to illegal content and to receive relevant results is the main underlying driver of risk of harm associated with search services. For the purposes of understanding the risks of encountering illegal content, a key difference between the different types of search services is the source of the content indexed and presented to users in the search results. In particular:

- **General search services (including downstream general search services)** present users with access to indexed webpages from across the entire clear web, while vertical search

[2204] Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T. and Fritz, M., 2023. Not what you've signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. [accessed 22 September 2023].
[2205] Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T. and Fritz, M., 2023.
[2206] For further information on the characteristics, see chapter 6: Introduction to the Register.

services have a far narrower scope, using an API (or equivalent technical means) to present content to users from pre-determined locations on the web.

- **Vertical search services** are likely to have a materially lower level of risk of harm than general search services. This is because vertical search services typically only focus on a specific segment of online content (such as particular products or services) and draw results via an API (or equivalent technical means) from pre-determined websites that may contain professional or curated content, rather than indexing sites from across the clear web. For example, a travel search site may be much less likely to present illegal content to a user, as the search feature on the site will be limited to hotels/flights/car rentals on the websites/databases of travel agents. This limited search functionality suggests that the risk of vertical search services providing access to illegal content is lower than with general search services.[2207]

General search services

24.24 The evidence suggests that using general search services is an effective way for users to access some types of illegal content when proactively trying to find it online. Although the total volume of illegal content directly accessible via general search services is unknown, the evidence presented here covers a broad range of content that may amount to an offence. It suggests that a user who knows what to look for can access a wide range of illegal content from among the billions of indexed web pages accessible via general search services. The examples below are those for which there is published evidence.

24.25 There is evidence that general search services can be used to access content that amounts to offences related to terrorism.[2208] Studies also show that these search services have returned content that relates to hate offences in relation to certain groups, notably antisemitic content.[2209]

24.26 Evidence suggests that general search services are used to access extreme pornography, with users searching for relevant terms that describe extreme pornographic content which would be prohibited.[2210] There is also evidence that searches for extreme pornography allow users to be led to videos directly from the results page.[2211]

[2207] At the time of writing, we are unaware of any clear web vertical search services that draw their search result content from databases of illegal content.

[2208] A 2022 report by Tech Against Terrorism highlights that 198 websites identified as being operated by terrorist actors existed on the surface web and were "*often easily discoverable through search engines*". Source: Tech Against Terrorism, 2022. The Threat of Terrorist and Violent Extremist Operated-Websites. [accessed 22 September 2023].

[2209] A study commissioned by the Antisemitism Policy Trust and CST tested the SafeSearch function on Google Images searches for two search terms. Antisemitism Policy Trust, 2021. Unsafe Search: Why Google's SafeSearch function is not fit for purpose. [accessed 26 September 2023].

[2210] From a dataset of 13,888 search instances from nearly 2,000 users looking for adult content, Ofcom found that 158 of these search instances from 40 users had a search term associated with extreme pornography before visiting an adult content site. This suggests that these search terms lead to content that at the very least contains a relevant matching description, and at worst is illegal extreme pornographic content. Source: Ofcom analysis of Ipsos Iris Clickstream Data, 15th September – 15th October 2021, UK, ages 15+.

[2211] Analysing a data set of 9,078 searches that led users to ten of the most popular adult content sites, one in five (1,831) of these searches led directly to a video. 53 (0.6%) instances of these searches included terms associated with extreme pornography. Source: Ofcom analysis of Ipsos Iris Clickstream Data, 15th September – 15th October 2021, UK, ages 15+.

24.27 General search services are identified as one of the most common methods of finding CSAM (child sexual abuse material) online,[2212] alongside U2U services. A study on websites hosting CSAM content found that websites generally do not hide their intention, such that *"if an individual can access and use a search engine with a modicum of skill, they can assuredly find [CSAM]".*[2213] Offenders have also confirmed their use of search engines in accessing CSAM[2214], and research has shown that major search engines have, in the past, provided direct access to illegal child abuse imagery from search requests.[2215]

24.28 There is also evidence showing the ways in which general search services can be used to access websites offering for sale or supply illegal items, such as drugs[2216], firearms[2217] or articles used in fraud offences such as stolen credit card details.[2218]

24.29 Search results have been shown to present users with content related to suicide and self-harm, including potentially illegal content.[2219] Studies on self-harm patients and male

[2212] Steel, C. M. S., 2015. Web-based child pornography: The global impact of deterrence efforts and its consumption on mobile platforms, *Child Abuse & Neglect.* [accessed 26 September 2023].
[2213] Westlake, B.G., Bouchard, M. and Girodat, A., 2017. How Obvious Is It? The Content of Child Sexual Exploitation Websites, *Deviant Behavior*, 38(3), pp. 282-293. [accessed 26 September 2023].
[2214] In a qualitative study on the pathways for accessing CSAM online in which interviews were conducted with 20 people who had viewed CSAM online and had been investigated by law enforcement. 2 respondents reported their initial exposure occurred via intentional searches on search engines. When asked about access methods generally, 13 respondents reported using search engines as a pathway to access CSAM. Note, we understand the data used in this study is from 2015. Source: Bailey, A., Allen, L., Stevens, E., Dervley, R., Findlater, D., and Wefers, S., 2022. Pathways and Prevention for Indecent Images of Children Offending: A Qualitative Study, Sexual Offending: Theory, Research, and Prevention, 17. [accessed 07 August 2024]
[2215] Researchers for AntiToxin surfaced CSAM on Bing from a variety of search queries, including overt and less obvious queries. In response to the findings Microsoft made changes to Bing to remove the identified content and block some of the search terms used. Source: TechCrunch (Constine, J.), 2019. 'Microsoft Bing not only shows child sexual abuse, it suggests it'. [accessed 07 August 2024].
[2216] Commission On Combating Synthetic Opioid Trafficking, 2022. Technical Appendixes. [accessed 17 October 2022].
[2217] Research has found that searches for stun guns, handguns and realistic imitation firearms returned results which appeared to be offers to supply prohibited weapons. Research has also shown that prohibited knives could be found on online marketplaces, which if the pages were indexed – which is highly likely – would also be accessible directly from search engines. Source: Which, 2022. Illegal weapons found for sale on Amazon, eBay, Wish and AliExpress.[accessed 26 September 2023]; Ofcom, 2023. Sale of prohibited items on search services. [accessed 26 September 2023].
[2218] Researchers found numerous search results and webpages offering to supply stolen credit card details to buyers, as well as guidance on how to commit fraud using this kind of information. Source: Ofcom, 2023. Articles and items for use in the commission of fraud – accessibility via search services. [accessed 26 September 2023].
[2219] A 2021 study investigating how search engines handle suicide search queries examined the top 20 search results returned in response to queries related to suicide. The study found that 22% of Microsoft Bing URLs, 19% of DuckDuckGo URLs and 7% of Google Search URLs were "harmful", that is, assessed by researchers to encourage, promote or facilitate suicide, or contain discussions of suicide methods. The researchers also looked specifically at search results encouraging suicide and found that this was the case for 10% of Microsoft Bing URLs, 8% of DuckDuckGo URLs and 4% of Google Search URLs. Source: Borge, O., Cosgrove, V., Cryst, E., Grossman, S., Perkins, S., and Van Meter, A., 2021. How Search Engines Handle Suicide Queries. Journal of Online Trust and Safety, 1(1). [accessed 07 August 2024]; Research commissioned by Ofcom and conducted by the Network Contagion Research Institute exploring the accessibility of content promoting self-injurious behaviour via search services identified 1% of the search results assessed as 'extreme', referring to "posts that encourage others to engage in self-injurious behaviour. This included treating self-injurious behaviour as a game; explicit invitations to join in self-injurious behaviour activities; and apps or other interactive spaces making it easy for people to engage in self-injurious behaviour." Note that the search queries tested were formulated with the intention that they would return such content if it was accessible via a search engine. Source: Ofcom, 2024. One Click Away: A Study on the Prevalence of Non-Suicidal Self Injury, Suicide, and Eating Disorder Content Accessible by Search Engines, [accessed 07 August 2024].

suicide victims found that both these groups use search engines to access pro-suicide content and to research methods.[2220]

24.30 Nonconsensual deepfake intimate images have also been shown to be easily accessible via the largest general search services.[2221] The tools that can be used to create this content, and websites that host large volumes of it, were also appearing high up in search results and even when search queries were not explicitly targeted at surfacing such deepfake content.[2222]

24.31 General search services (including downstream search services) have also been identified as having a potential risk of foreign interference, such as from those who seek to manipulate information to advance a particular agenda. This includes from those working for or on behalf of foreign governments who can attempt to leverage search results with the intention to benefit a foreign power. This can include by manipulating public discourse, sowing discord, discrediting the political system and undermining the safety or interests of the UK. Research has demonstrated that these risks, can appear in search results presented to users, although the volume of content and likelihood of it amounting to the relevant offence varies significantly across search services, language used and search queries.[2223]

24.32 A critical component of general search services (including downstream search services) is their ability to present users with the most relevant content based on their search query. The search engines that underpin the operation of general search services use proprietary algorithms ('ranking') to perform this prioritisation function. The ranking process uses factors such as how closely the search query is matched and the website's functionality and authority (the perceived value of the site's content and how often it is linked to by other sites). As with all functionalities, the ranking process is designed to provide accurate and reliable content, but it can be manipulated to increase the likelihood of illegal content being displayed to users. For example, the tactic of keyword stuffing (filling a web page with

---

[2220] A study involving self-harm patients admitted to hospital has found that this group would often go online to 'research' methods with the intention of planning an effective attempt. The type of content this group would view would include pro-suicide content, including consulting medical, academic and other 'factual' resources. Similarly, the most common behaviour among a study of 288 men aged 40-54 who had committed suicide was searching for information on suicide methods (10%). Of this 10%, a third (33%) had died by the method they were known to have had searched about. Source: Biddle, L., Derges, J., Goldsmith, C., Donovan, J.L. and Gunnell, D., 2018. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital, *PLoS One, 13(5);* The National Confidential Inquiry into Suicide and Safety in Mental Health (NCISH), 2021. Suicide by middle-aged men. [accessed 26 September 2023].

[2221] Journalists found that pornographic images featuring likenesses of female celebrities were the first images that Google and Bing surfaced in response to search queries including female celebrities names and words such as "deepfake", "deepfake porn" or "fake nudes". Source: NBC News (Tenbarge, K.), 2024. Google and Bing put nonconsensual deepfake porn at the top of some search results. [accessed 26 September 2024]

[2222] My Image My Choice analysed deepfake abuse content accessible via Google search. On the top 40 websites dedicated to deepfake abuse they analysed, they identified over 270,000 videos which had gained over 4 billion views. Google Search appeared to be was driving 68% of traffic to these sites. They also found deepfake abuse content was returned even when using search queries not explicitly searching for it – with deepfake abuse content often making up the top and majority of the results on the page returned. Source: My Image My Choice, 2024. Deepfake Abuse: Landscape Analysis 2023-24. [accessed 14 August 2024]

[2223] Research conducted by the Alliance for Securing Democracy, on behalf of Ofcom, focused on identifying the presence and prevalence of state-backed media outlets in search results pages for a range of general search services. It is important to note that the goal was to assess the potential risk of foreign interference by identifying types of variables – such as the topic searched for and language used – that regularly generated search results from state-linked websites, and not to assess whether the content of those search results would be illegal under the UK's Foreign Interference Offence (National Security Act 2023). Source: Schafer, B. & Benzoni, P (*Alliance for Securing Democracy, on behalf of Ofcom*), 2023. Assessing the Risk of Foreign Influence in UK Search Results. [accessed 18 October 2024].

keywords or numbers in an attempt to manipulate rankings in search results)[2224] has been identified in research looking at how easily illegal content relating to fraud can be accessed via search services.[2225]

24.33 Although there is not necessarily published evidence relating to the risk of encountering content via search services that amounts to every specific offence covered by the OSA, we are of the view that this does not equate to a lack of risk in relation to these harms. For the reasons set out above and where there is evidence for certain content types – or adjacent content types – existing online, including in publicly available spaces on user-to-user services or stored on file-sharing services, we think it is reasonable to assume that such content could be findable via search services. For example, potentially obscene torture content, or links ostensibly directing users to user-to-user sites with this content (including those for sharing 'gore' images and videos) are easily discoverable via results from search services. In this case, we believe it is reasonable to assume that human or animal torture content, or potentially content that amounts to the animal cruelty offence could be accessed through search.

Risk factors: User base

24.34 The user base includes the size and composition of the users of a search service, covering demographics and other characteristics. Although the user base is included here as a characteristic, it is only considered in a very limited way, compared to U2U services' user bases. This is because user bases on search services are particularly difficult to measure, as in most cases there is no need to have an account to be able to use the service.

User base size

24.35 The size of a search service's user base is considered a risk factor and, therefore, something that needs to be considered within a service provider's risk assessment. All things considered, the larger a user base, the more people who have the potential to encounter illegal content or activity – intentionally or otherwise – via that service if other mitigating factors have not removed that risk.

User base demographics

24.36 While anyone using the internet is likely to use search services to some degree, it is reasonable to assume that user base demographics will differ from one service to the next.

24.37 As evidenced throughout the Part 1: User-to-user services chapters, certain kinds of illegal harm are more prevalent among certain groups. Therefore, the demographic characteristics of a search service's users should be considered as a factor that influences the relative risk of harm occurring via that service. For example, vulnerable users (and particularly those with multiple protected characteristics) could be impacted differently from harm that they may encounter in search results.

[2224] Google, n.d. Spam policies for Google web search. [accessed 26 September 2023].
[2225] Ofcom, 2023. Articles and items for use in the commission of fraud – accessibility via search services. [accessed 26 September 2023].

Risk factors: functionalities

24.38 These include the underlying potential for illegal content on webpages indexed by general search services to appear in, or via, search results; the features visible to users to optimise search results (such as recommended searches, autocomplete suggestions); and those which determine results behind the scenes (such as ranking algorithms).

24.39 These service characteristics are designed largely to optimise the accuracy and usefulness of search results to users. Where a user is intentionally seeking out illegal content – which is considered the most likely situation in which a user would encounter content that amounts to an offence – these same optimising characteristics have the unintended consequence of helping that user encounter illegal content.

Search query inputs

24.40 Functionalities enabling users to input search queries can impact what search queries are made and may therefore influence the results that are presented to users. As set out previously, simply entering search queries aimed at sourcing illegal content can be an effective means of finding it. While targeted search queries, such as those using slang or coded terminology tend to be more effective, the evidence also highlights that innocuous or non-specialist search queries also returned potentially illegal content. The ease with which some content can be found is compounded by a concern that those who are actively searching for such content may be more susceptible to experiencing or causing harm as a result.[2226]

Image search

24.41 Specifically, the ability to use images as a query to find other images or relevant results ('reverse image search') has been demonstrated as an effective way to find services selling drugs via general search services.[2227] While published evidence for other buying/selling offences is limited, it is possible that the reverse image search functionality also presents opportunities to access content relating to other prohibited items.

24.42 It has also been noted that the ability to conduct reverse image search can be misused by those seeking to identify and potentially locate individuals. This ability to search for new or recent information online about someone based on their image has been flagged as a serious concern in relation to offences related to harassment, coercive and controlling behaviour and intimate image abuse.[2228]

Search prediction and personalisation

24.43 Functionalities that make suggestions related to a user's search requests can help users be more targeted or accurate in their searches and have also been shown to influence users' search strategies in relation to a range of illegal content. It is reasonable to assume that these functionalities can increase the risk of accessing illegal content amounting to a range

[2226] For instance, research exploring the impact of exposure to potentially radicalizing information suggests that individuals who actively seek out terrorism content are at a higher risk of radicalisation. Source: Schuman, S., Clemmow, C., Rottweiler, B., Gill, P. 2024. Distinct patterns of incidental exposure to and active selection of radicalizing information indicate varying levels of support for violent extremism. [accessed 30 August 2024].
[2227] RAND, 2022. Commission On Combating Synthetic Opioid Trafficking. [accessed 26 September 2023].
[2228] We recommend readers refer to the related chapters in the U2U section of this volume.

of offences via general search services (including downstream search services), unless effective mitigations are in place to prevent this, or indexed content is blocked.

24.44 For content linked to suicide or self-harm, evidence has shown that the predictive element of a search bar can suggest potential methods or instructions on how to self-harm or end one's life.[2229] Evidence has also found that search bar predictions have recommended hateful or racist search queries.[2230]

24.45 Other research has also demonstrated how recommended or suggested searches and autocomplete functions have pointed users to fraud-related content such as stolen credit card details[2231], and in the past even enabled users to find child sexual abuse imagery.[2232]

Risk factors: Business models and commercial profiles

Revenue models

24.46 General search services typically generate revenue using an advertising-based model.[2233] Search services are paid by users and/or businesses to display ads for their products/services alongside search results. For example, advertisers may pay the search service whenever a user clicks on their advert or sponsored link. This is the main pricing structure used by Google.[2234] We understand that downstream general search services also earn revenue through advertising.[2235]

24.47 There is limited evidence on the links between different revenue models and the presence of illegal content in search results. Nevertheless, evidence suggests that advertisements on search services may be misused for illegal activity.

24.48 Indeed, advertisements on search services can suggest products and sites to users that may enable them to engage in illegal behaviours; for example, spyware (products which allow users to track and monitor other people's devices) is often cited as a facilitator in cases of

[2229] For more detail see the Register of Risks chapters 'Encouraging or assisting suicide' and 'Encouraging or assisting serious self-harm'. Google search results as of 9th March 2022, examples provided to Ofcom by the Samaritans.
[2230] A study found that Google recommended *"queers should be shot"* when the first two words were typed into its search box (Google stopped recommending such phrases a week after these examples were flagged). Loeb, J., 2018. Google is 'promoting hate speech', claims internet law expert, *E&T*, 22 January. [accessed 27 September 2023].; Similarly, the Antisemitism Policy Trust reported that Microsoft Bing directed users to hateful searches with the autocomplete *"Jews are bastards*". Antisemitism Policy Trust response to 2022 Ofcom Call for Evidence: First phase of online safety regulation; Wired (Lapowsky, I), 2018. Google Autocomplete Still Makes Vile Suggestions. [accessed 18 November 2024]. Please note some of these sources are from 2018 and search services have made changes to their systems since then, but this highlights the risks present when mitigating measures are not in place or up to date.
[2231] Ofcom, 2023. Articles and items for use in the commission of fraud – accessibility via search services. [accessed 26 September 2023].
[2232] In 2019, researchers for AntiToxin who showed that CSAM could be accessed with relative ease on Bing Search, also highlighted that some auto-complete suggestions led to illegal content, including from innocuous search queries not initially constructed to intentionally surface CSAM. In response to the findings Microsoft made changes to Bing to remove the identified content and block some of the search terms used. Source: Constine, J., 2019. 'Microsoft Bing not only shows child sexual abuse, it suggests it', TechCrunch, 10 January. [accessed 07 August 2024].
[2233] Some search engines use a subscription model in lieu of advertising to generate revenue, although this is rare.
[2234] Competition and Markets Authority, 2020. Online platforms and digital advertising. [accessed 26 September 2023].
[2235] Competition and Markets Authority, 2020; Australian Competition and Consumer Commission, 2021. Digital platform services inquiry. [accessed 26 September 2023].

coercive control.[2236] Although these devices are illegal, the research suggests that they are sometimes marketed as parental safeguarding tools.[2237]

24.49 There is also some evidence to suggest that foreign interference campaigns manifest on search services. These are primarily related to paid advertisements. The tactics used include the manipulation of search engine optimisation techniques, including platform advertising mechanisms, and the use of state-owned services to obscure the truth.

24.50 For example, Google AdWords (an advertising system targeting key search terms) was used in Russia to manipulate Google search results during the US 2016 Presidential election.[2238] This was accomplished by purchasing search ads on Google attacking political candidates participating in the election. Google, in its Threat Analysis Group (TAG) bulletins, documents how it terminated Ad Accounts associated with foreign interference. For example, in April 2022 Google terminated nine AdWords accounts as part of its *"investigation into coordinated influence operations linked to Russia. The campaign was linked to the Internet Research Agency (IRA) and was sharing content in Russian, French, Arabic, and Chinese that was supportive of Russia's 2014 invasion of Crimea and the Wagner Group's activity in Ukraine and Africa*".[2239]

Growth strategy

24.51 Developing an expansive search index is a core component in a general search service's growth strategy. A provider's approach to expanding its index may influence the risk of harm. If providers do not carry out due diligence when indexing web pages, there is a risk that illegal content will be indexed and thereby become accessible by users of the service. This issue is likely to come into play if a provider is looking to develop its own index rapidly in order to compete with more established services, prioritising this over user safety.

Commercial profile

24.52 Despite the limited evidence, we consider that search services that are low-capacity or at an early stage in their lifecycle may face an increased risk of harm on their services.[2240]

24.53 Low capacity and early-stage services often have limited ability to develop and deploy targeted mitigations or moderation measures to reduce the risk of users encountering illegal content in or via search results. For instance, they may have limited technical skills, financial resources or lack access to data from useful third parties (for example, CSAM URL lists from the IWF).

24.54 We understand that when a downstream search service syndicates some or all of its search results from an upstream supplier, some of the safety measures applied on the upstream service may be extended (in whole or in part) to the downstream service, depending on the particular syndication arrangement in place between the entities. Therefore, we recognise

---

[2236] Sugiura, L., Blackbourn, D., Button, M., Hawkins, C., Tapley, J., Frederick, B., Nurse, J. R. C. and Belen-Salam, R., 2021. Computer Misuse as a Facilitator of Domestic Abuse. [accessed 26 September 2023].
[2237] Sugiura, L., Blackbourn, D., Button, M., Hawkins, C., Tapley, J., Frederick, B., Nurse, J. R. C. and Belen-Salam, R., 2021.
[2238] Bradshaw, S, 2019, *Disinformation Optimised: Gaming Search Engine Algorithms to Amplify Junk News*, Internet Policy Review, 8(4). [accessed 26 September 2023].
[2239] Google Threat Analysis Group (Huntley, S.), 2022. TAG Bulletin: Q2 2022. [accessed 26 September 2023].
[2240] We have analysed evidence of how services with low capacity, or at early stage, may increase the risk of certain illegal harms on U2U services. We consider that the same reasoning broadly applies for all harms on U2U services (see Part 1: User-to-user services chapters) and also applies to search services.

that the risk profile of a low capacity, early-stage service, that is also a downstream search service, may vary from that described at 24.53. It will be the responsibility of the provider of the relevant search service to assess any factors that impact the risk profile of that service.

24.55 We discuss our position as to what obligations fall on downstream search services under what circumstances in ‘ Our approach to developing Codes measures’.

25. Governance, systems and processes

Introduction

25.1 As part of this risk assessment for illegal content, we have assessed the characteristics of a service relating to governance, systems and processes separately. This is because the analysis of risk arising from these characteristics applies to different kinds of illegal harms.

25.2 This chapter assesses how the governance structures, systems and processes for a service may be relevant to the risk of user exposure to illegal harm on that service. We expect service providers to use this information when considering how their own current or future governance structures, systems and processes will affect user safety.

25.3 We recognise that governance structures, systems and processes are often put in place by service providers to mitigate the risk of harm to users.[2241] However, service providers should consider how they are implemented, as this can also affect wider online safety decisions and responsibilities. In fact, if implemented ineffectively, governance structures, systems and processes can have the inverse effect of increasing the risk to users in some contexts.

25.4 After reviewing available evidence, we have identified two general scenarios where risk can arise from these areas themselves: (a) governance and/or other systems and processes currently in place within regulated services that are inadequate; and/or subsequently (b) an absence of such governance and other systems and processes. This chapter provides service providers with additional information on these risks.

Definitions

25.5 The Online Safety Act (the Act) does not define 'Governance'. Based on our understanding of the sector, and the evidence consulted, Ofcom has interpreted the concept of governance in the context of online safety as "any structure, or structures to ensure that decisions are made with adequate oversight, accountability, transparency and regard to online safety compliance, specifically in relation to risk management, product and content governance within a service".[2242]

25.6 On the other hand, 'systems and processes' are described in the Act as "any human or automated systems and/or processes, and accordingly, includes technologies".[2243] For the

[2241] For example, content moderation systems can play a crucial role for services to detect and remove illegal content swiftly, such as in cases of terrorist acts being livestreamed on a U2U service.
[2242] Milliman, 2021. Report on principles-based best practices for online safety Governance and Risk Management. This report was commissioned by Ofcom. This definition aligns with Milliman's description of governance as made up of the concepts of individual and overall accountability, non-executive oversight, independent executive oversight, oversight of risk strategy and appetite, monitoring of the effectiveness of risk management, effective communication of risk and setting an appropriate risk culture and aligned incentives. We consider that in the context of online safety, governance relates more broadly to structures which work to ensure that decisions are aligned with user safety at all levels of an organisation.
[2243] Section 236 of the Online Safety Act.

purposes of this risk assessment, we interpret this to mean any series of actions taken by a service provider, including actions that mitigate the risks of harm arising from illegal content being encountered that may not have been addressed elsewhere in the Register of Risks.

Evidence of risks of harm to individuals arising from governance, systems and processes

25.7 We structure this section as per our Codes of Practice (Volume 2) with the addition of Governance which can be found in Volume 1, Chapter 5.

a) Governance (U2U and Search)
b) Moderation (U2U and Search)
c) Search design (Search)
d) Reporting and complaints (U2U and Search)
e) Recommender systems (U2U)
f) Service design and user support (U2U)
g) Terms of service and publicly available statements (U2U and Search)
h) User access (U2U)

Governance (U2U and Search)

25.8 Harm can arise from inadequate or absent governance arrangements in several ways. We have drawn on risk management practice and case studies from other sectors, including financial services, health and safety and manufacturing, to understand safety failures due to inadequate and insufficient governance arrangements. We have also consulted the literature on corporate governance and best practice to inform our understanding of risk.

25.9 Our review of good practice standards and principles in risk management and corporate governance across a range of different industries demonstrates the importance of clear, consistent, and codified assurance processes, governance structures, reporting mechanisms and internal communications in ensuring good safety practices and positive outcomes for users and consumers. Further information as to how service providers can mitigate the risks outlined here can be found in our Codes of Practice (Volume 2).

Governance arrangements

25.10 Users may be more likely to be exposed to illegal content where there is insufficient oversight and scrutiny of risk management activities. One of the key remits of a governance body is to consider the effectiveness of a company's risk and governance practices in its decision making.[2244] Evidence demonstrates that the structure of a governance body influences organisational approaches to risk management; for instance, it has been found that the presence of non-executive directors in the boards of financial institutions is linked

[2244] OECD, 2023. G20/OECD Principles of Corporate Governance. [accessed 11 November 2024]; Milliman, 2021. Report on principles-based best practices for online safety Governance and Risk Management. [accessed 11 November 2024].

with a more risk-averse attitude towards investments, as non-executive directors may be more concerned about their reputations.[2245]

25.11 Evidence shows that where there is a failure in a governance body to fulfil its functions, risk management activities may not be adequately challenged or scrutinised.[2246] This could result in lack of oversight across a range of activities, including the implementation of appropriate and effective mitigations, the changing level and type of risks of harm to users, and broader organisational compliance with safety outcomes.[2247] For example, analysis of a high-profile safety incident in the aerospace industry showed that the failure of a board to account for safety risks contributed to fatality events, as there were no mechanisms to escalate safety concerns to a board or oversight body.[2248]

Senior accountability and responsibility

25.12 Our analysis found that senior accountability for online safety is critical in building a culture that prioritises safety for users. In this context, senior accountability refers to a structure where senior members of staff are expected to answer for user safety decisions, and to own responsibility for decision-making. As part of this, individual accountability is also particularly important – it was the first principle of good governance set out by Milliman in their work looking at principles-based best practice for online safety governance and risk management.[2249]

25.13 Users may be more likely to be exposed to illegal content if there is a lack of accountability for compliance at senior management level. This is because it may remove senior oversight and responsibility for user safety decisions and fail to oversee and address risk management activities by an overall governance body or board.

25.14 The importance of senior accountability for mitigating risks is recognised in other sectors, such as banking, health and safety[2250] and AI governance. A study evaluating a regulation designed to address ongoing risk management failures in Australia's banking sector found that *"greater felt accountability among senior executives stimulates more proactive and diligent risk management behaviour"*. It anticipated that when accountability cannot be delegated, bad outcomes reflect badly on the accountable executives themselves, so they should be less likely to ignore red flags and instead seek out more risk information and

---

[2245] Akbar, S., Kharabsheh, B., Poletti Hughes, J. and Shah, SZA., 2017. Board Structure and Corporate Risk Taking in the UK Financial Sector, *International Review of Financial Analysis*, Volume 50, pp. 101-110. [accessed 11 November 2024].

[2246] The Health and Safety Executive offers several case studies of negative safety consequences when board members do not lead effectively on health and safety management. Source: Health and Safety Executive, n.d. Case studies: When leadership falls short. [accessed 20 November 2024].

[2247] An analysis of past incidents [including major safety incidents in high hazard industries] by the OECD indicates that inadequate leadership have been a recurrent feature, *"including the monitoring of safety performance indicators at Board level"*. Source: The Organisation for Economic Co-operation and Development, 2012. Corporate Governance for Process Safety: OECD Guidance for Senior Leaders in High Hazard Industries. [accessed 19 September 2023].

[2248] This includes lawsuits filed against Boeing following the crashes of two 737 MAX airplanes in 2018 and 2019, in which shareholders claimed that a failure of the board to account for safety risks contributed to fatality events: *"safety was no longer a subject of Board discussion, and there was no mechanism within Boeing by which safety concerns… were elevated to the Board or to any Board committee."* Source: The Washington Post, 2021. Verified Amended Consolidated Complaint. [accessed 5 May 2023].

[2249] The importance of individual senior accountability is stressed within the Three Lines Defence model that has been implemented across many sectors and was drawn on Milliman in their analysis. Source: Milliman, 2021. Report on principles-based best practices for online safety Governance and Risk Management. p.18.

[2250] Health and Safety Executive, 2013. Leading health and safety at work: Actions for directors, board members, business owners and organisations of all sizes. [accessed 24 August 2023].

evaluate it more carefully.[2251] Evaluation of a similar regime in the UK banking sector, introduced to hold senior management to account for failures that occurred on their watch, found that the majority of senior managers and firms which reported these had brought about positive and meaningful changes to behaviour in the industry.[2252] Furthermore, senior leadership failures in financial services in relation to the 2008 financial crisis are taken as a case study in in the risks of reduced oversight and subsequent excessive risk-taking.[2253]

25.15 The ICO's guidance about AI risk management regarding data protection states that senior management are accountable for addressing the technical complexities of AI and cannot delegate this responsibility to others. It states that senior management will need to align internal structures, roles and responsibilities maps, training requirements, policies and incentives to its overall AI governance and risk management strategy.[2254] Likewise, the AI Risk Management Framework by the US National Institute of Standards and Technology (NIST) states that effective risk management is realised through organisational commitment at senior levels and may require cultural change within an organisation or industry. It says that organisations need to establish and maintain the appropriate accountability mechanisms, roles and responsibilities, culture, and incentive structures for risk management to be effective.[2255]

Internal assurance and compliance functions

25.16 Evidence indicates that a governance framework with strong internal controls leads to effective risk management. Respondents to a European Commission consultation agreed on the need to improve governance and were supportive of "strengthening, clarifying and harmonising responsibilities of board members; ensuring effective risk management and internal controls."[2256] In the context of online safety, inappropriate risk mitigation and management evaluation processes can lead to users being exposed to illegal content. These risks may also arise where such processes are inconsistent, where measures are ineffective at addressing specific risks, or where measures are not future-proof.[2257] For these reasons, internal assurance and compliance functions can be effective in ensuring there is adequate oversight over risk management.

[2251] Note that this study has a relatively small sample size of 41 interviews with accountable persons. Source: Elizabeth Sheedy and Dominic Canestrari-Soh, 2023. Does executive accountability enhance risk management and risk culture?, Accounting & Finance, 63(4). [accessed 17 April 2024].

[2252] The Senior Managers and Certification Regime (SM&CR) for banks and insurers, launched in 2016 in the UK, requires the most senior decision-makers in firms to have clearly assigned responsibilities, and to be accountable for actions within their remit. Results from a survey of banks and insurers showed that 94% of senior managers and 96% of firms which responded reported that the SM&CR had brought about positive and meaningful changes to behaviour in industry. Source: Bank of England, 2020. Evaluation of the Senior Managers and Certification Regime. [accessed 17 April 2024].

[2253] Additionally, in the aftermath of the 2008 financial crisis, an inquiry into professional standards and culture of the banking sector by the Parliamentary Commission on Banking Standards concluded that many bankers had been allowed to operate with little accountability, and *"claimed ignorance or hid behind collective decision-making"*. Financial Conduct Authority, 2013. The FCA's response to the Parliamentary Commission on Banking Standards. [accessed 3 May 2023].

[2254] ICO, What are the accountability and governance implications of AI?. [accessed 17 April 2024].

[2255] US National Institute of Standards and Technology (NIST), 2023. Artificial Intelligence Risk Management Framework. [accessed 17 April 2024].

[2256] European Commission, 2022. Public consultation on the strengthening of the quality of corporate reporting and its enforcement: summary report. [accessed 03 May 2023].

[2257] A report by Ofcom on the Buffalo attack concluded that services should make efforts in product and engineering design processes to prevent the upload of terrorist content in an effort to prevent similar incidents in the future. Ofcom, 2022. The Buffalo Attack: Implications for Online Safety. [accessed 4 October 2023].

25.17 Evidence from examples of high-profile organisational failures highlight the importance of effective internal controls in managing and mitigating a range of risks.[2258] Our analysis points to weak or absent controls as a key contributory factor to organisational failure, especially in major corporate scandals.

25.18 We found evidence supporting the hypothesis that poor internal controls played a role in high-profile instances of organisational failures related to fraud,[2259] data integrity[2260] and product safety[2261] across other sectors.

Staff incentives, policies and processes

25.19 Our analysis suggests that there is an increased risk of harms to users if staff across the service provider are not adequately trained on compliance.[2262] This can lead to a weaker culture of risk mitigation and management across the organisation and may in turn result in the inconsistent application of risk mitigation and management measures.

25.20 Without efforts to align safety objectives across different areas of a service provider, it is possible that staff will not understand how the provider is approaching regulatory compliance, or how it manages and mitigates risks of illegal content being displayed to users on its service(s). This is supported by evidence of how the absence of compliance training programmes has contributed to serious corporate scandals.[2263]

[2258] Di Miceli Da Silveira, A., 2011. 'Corporate Scandals of the Earlier 21st Century: Have We Learned the Lessons?' [accessed 03 May 2023].

[2259] This includes the case study of petrochemical operators Petrobras and PdVSA, where systematic violation of internal controls and the absence of controls in key areas led to a failure to prevent or mitigate fraudulent activity. Source: Hamilton, S. and Micklethwait, A., 2006. *Greed and corporate failure: The lessons from recent disasters*. Springer; Omoteso, K., Obalola, M., 2014 'The Role of Auditing in the Management of Corporate Fraud' in Said, R., Crowther, D., Amran, A. (eds.) *Ethics, Governance and Corporate Crime: Challenges and Consequences*, Emerald Group Publishing Limited, pp.129-151.; Burger, M., Taken Smith, K., Murphy Smith, L. and Wood, J., 2022. An examination of fraud risk at oil and gas companies, *Journal of Forensic and Investigative Accounting*, pp.74 – 85.

[2260] 2017, the FDA sent a warning letter to Indian pharmaceutical company Wockhardt, warning of repeated failures in oversight and controls that had contributed to the deletion of data related to failed tests. US Food & Drug Administration, 2017. Warning Letter, Morton Grove Pharmaceuticals, Inc. [accessed 4 October 2023].

[2261] *"In the absence of any focus or controls on airplane safety, the Boeing Board pushed for achievement of production deadlines and competition with its chief rival, Airbus. In reviewing and approving the 737 MAX project, the Board never examined, considered, or questioned potential safety issues resulting from the re-design of the earlier generation 737 NG."* Source: Volkov Law Group, 2021. Boeing's Board Governance Failures and the 737 MAX Safety Scandal (Part III of IV). [accessed 4 October 2023].

[2262] In the case of Siemens, which in 2008 was subject to regulatory investigations for bribery, the failure to embed a programme of compliance and Code of conduct for staff has been cited as playing a "decisive role" in the scandal. Source: Primbs, M. and Wang, C., 2016. Notable Governance Failures: Enron, Siemens and Beyond. Comparative Corporate Governance and Financial Regulation. 3. [accessed 21 September 2023].

[2263] Primbs, M. and Wang, C., 2016.

Moderation (U2U and Search)

25.21 Moderation, whether automated,[2264] [2265] human, or a combination of both, is put in place by service providers to identify and take action on content that is illegal or does not meet the content policies of the service in question. Here we use the terms 'content moderation' to refer to moderation on U2U services and 'search moderation' to refer to moderation on search services, or 'moderation' if referring to both.

25.22 Moderation systems which are poorly designed, deployed or resourced may fail to mitigate or minimise the risk of users being exposed to illegal content online. The risks of harm associated with these are outlined in each chapter in the Register of Risks. Further information as to how service providers can implement content or search moderation effectively and mitigate the risks described here can be found in the Codes of Practice (Volume 2).

Ineffective moderation

25.23 Ineffective moderation strategy, systems, processes and technology can lead to the risk of users being exposed to illegal content. Online services where providers deploy less stringent moderation, for whatever reason, may be viewed as preferred spaces for the dissemination of content that would amount to a wide range of offences compared to services with more strict or consistent moderation.[2266] This is part of an observed trend, where services with less stringent moderation, which often correlates with having fewer staff, have, on average, had higher volumes of terrorism content than services with stringent moderation.[2267] A report by the Molly Rose Foundation also found that "inconsistent and at times erratic" content moderation undermined the harm reduction strategies of three popular services.[2268] It has also been found that the relatively minimal moderation of a messaging service, allowing users to create groups and channels, had "lowered the hurdle for engaging

[2264] Automated moderation technology can support the identification and removal of priority illegal content, either when it is uploaded or once it is on a service. This includes tools that can compare each piece of content against a database or list of known illegal content using methods such as hash-matching, URL detection, or text detection. Any content that matches existing content in such a list or database can then be flagged for further review or automatically removed. This type of moderation can be highly effective to identify and remove specific types of illegal material, particularly due to its advantages at scale. Due to the sheer volume of content that may be available on a service (particularly on larger services) human moderation often benefits from assistance from automated processes.

[2265] A note on our automated moderation measures in our Codes of Practice (Volume 2) at the time of publication: We recognise that automated moderation systems and processes may be used to address several types of illegal harms, including child sexual abuse material, fraud, and terrorism. However, at the time of this publication, our approach focuses on the use of automated moderation systems, for both user-to-user and search services, that operate by detecting matches for known child sexual abuse material (CSAM). In our measures we did consider proposing the use of automated moderation, specifically keyword detection, to address content containing articles of fraud. However, we are still considering the most appropriate type of automated moderation technology to use to address this type of illegal behaviour.

[2266] For example, research has identified that users sharing content potentially amounting to hate and terror offences – in this case, those on the 'extreme right' – have shown a preference for utilising online services perceived as having more limited content moderation as well as utilising multiple services in an attempt to evade content moderation efforts of any one particular service provider. Source: The Institute for Strategic Dialogue (O'Connor, C.), 2021. Gaming and Extremism: The Extreme Right on Twitch. [accessed 11 November 2024].

[2267] Tech Against Terrorism, 2023. TCAP Insights. Patterns of online terrorist exploitation. [accessed 20 November 2024].

[2268] Molly Rose Foundation, 2023. Preventable yet pervasive: The prevalence and characteristics of harmful content, including suicide and self-harm material, on Instagram, TikTok and Pinterest. [accessed 17 April 2024].

in the politics of hate and has enabled extremist networks to propagandise, network and organise", which could eventually result in individuals being exposed to radicalisation.[2269]

25.24 Employing effective moderation strategies is all the more important in light of evidence suggesting efforts by some users to evade or bypass content moderation on services,[2270] as discussed, for example, in our Register of Risks chapters: Encouraging and assisting serious self-harm, Drugs and psychoactive substances, and Search.

Resourcing and time constraints

25.25 Resource constraints on moderation teams could lead to illegal content remaining on a service for a longer time. An Ofcom report noted that in service providers' content moderation processes there is typically a time-lag between content being referred and it being reviewed, due to resource constraints, and the potentially large and fluctuating volume of potentially illegal or harmful content referred.[2271] A study has suggested that the reduction of staff working on content moderation for a large service led to a major increase in the quantity of antisemitic content on the service.[2272] [2273]

25.26 Time pressures on human moderators may increase the risk of human error in moderation decisions. A report by Demos highlighted that human content moderators have to make decisions in minutes, often about content in a language or a context they do not understand, making mistakes inevitable.[2274] Moderation can be a challenge simply due to the large volumes of content that need to be sifted through; a point publicly acknowledged by a large search service: "The breadth of information available online makes it impossible to give each piece of content an equal amount of attention, human review, and deliberation. Even if that were possible, reasonable people could disagree on appropriate outcomes".[2275]

25.27 Inadequate training, including on abuse that is targeted at women, minority ethnic groups and people with intersecting identities, can lead to an uneven application of moderation standards.[2276] In response to a Call for Evidence, Refuge stated that "*to the untrained eye,*

[2269] HOPE not hate and the Antisemitism Policy Trust, 2021. Antisemitism and Misogyny: Overlap and Interplay. [accessed 4 October 2023].

[2270] For example, users have been found to send certain emojis to indicate that they sell drugs in place of text on a social media service. Source: Moyle, L., Childs, A., Coomer, R. and Barrat, M.J., 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, *International Journal of Drug Policy*, 63, 101-110. [accessed 3 June 2023].

[2271] Ofcom, 2023. Content Moderation in user-to-user online services.

[2272] A report by CASM Technology and ISD found a major increase in the number of antisemitic posts, coinciding with a reduction in content moderation staff at one social media service, saying the analysis demonstrates *"the broader and longer-term impact that platforms de-prioritising content moderation can have on the spread of online hate."* Note: On its methodology, the report comments there are 'inherent challenges in training language models on as nuanced a topic as antisemitism, but this architecture is evaluated to operate with an accuracy of 76%. Source: CASM Technology and the Institute for Strategic Dialogue (ISD), 2023. Antisemitism on Twitter Before and After Elon Musk's Acquisition. [accessed 17 April 2024].

[2273] In late 2022, ADL noted an increase in antisemitic content on the same service and a decrease in the moderation of antisemitic posts. Source: The Anti-Defamation League, 2022. Extremists, Far Right Figures Exploit Recent Changes to Twitter. [accessed 4 October 2023].

[2274] Demos (Krasodomski-Jones, A.), 2020. Everything in Moderation: Platforms, communities and users in a healthy online environment. [accessed 17 April 2024].

[2275] Google, 2020. Information quality & content moderation. [accessed 20 November 2024].

[2276] The Glitch response to Ofcom's 2022 Call for Evidence states that without "*comprehensive training for moderators about online gender-based violence and different tactics of online abuse, and how abuse specifically targets women, Black*

tech abuse can often be hard to recognise without an understanding of the broader context of domestic abuse and coercive control."[2277] As part of our video-sharing platform (VSP) regulation, we said that although providers for most of the regulated services we examined did have training materials for content moderators in place, including definitions of prohibited content, it is also important to build moderators' awareness of the cultural, linguistic, historical, and political context in the UK, to help them protect UK users.[2278]

25.28 It is important to have processes in place to prepare moderation staff for times of crisis. As part of our VSP regulation, we found that only some VSP providers provided detailed guidance on what to do in a crisis situation. Our report on the 2022 Buffalo Attack highlighted the importance of VSP providers having appropriate moderation measures and internal processes in place to respond quickly to crisis events as they arise. We considered that moderators who have access to high-quality resources will be better equipped to identify harmful content quickly, consistently and accurately.[2279]

Search design (Search)

25.29 The way in which search services are designed can affect the risk of harm to individuals. Users may encounter illegal content via the results of their query on a search service, as explored in the 'Search' chapter. Further information on how search services can be designed to mitigate risk of harm can be found in the Codes of Practice (Volume 2).

Predictive search functionalities

25.30 Predictive search functionalities[2280] can increase the risk of individuals receiving search suggestions that direct them to illegal content. This is because a predictive search suggestion might prompt a user to search for illegal content that they might otherwise not have searched for had the query not been suggested. Respondents to our 2022 Call for Evidence suggested that predictive search functionalities such as autocomplete play a role in increasing the discoverability of harmful and potentially illegal content on search services; in particular, autocomplete search functionalities may point users to content that encourages self-harm,[2281] or to hateful and racist content that can lead eventually to illegal content.[2282] One respondent to our Call for Evidence stated that "*search services have been*

---

and minoritised communities and users with intersecting identities is paramount – without this moderation risks being ineffective, inequitable and/or discriminatory". Glitch is a UK charity which exists to end online abuse and to increase digital citizenship across all online users. The Antisemitism Policy Trust response to Ofcom's 2022 Call for Evidence says that without quality assurance and independent scrutiny of moderator training, it risks not being effective at responding to harm. The Antisemitism Policy Trust is a charity that works to educate and empower parliamentarians and policy makers to address antisemitism. In the NSPCC response to Ofcom's 2022 Call for Evidence it says that "*Some online service providers rely primarily on volunteers in their own communities to self-police, with administrators doing occasional checks. There is a real concern here that the subjective nature of this moderation process creates inconsistency and potential for gaps in protection of users*". The NSPCC is a UK charity with over 130 years in experience safeguarding children from harms.

[2277] Refuge response to Ofcom's 2022 Call for Evidence.

[2278] Ofcom, 2023. Regulating Video-Sharing Platforms (VSPs). Our first 2023 report: What we've learnt about VSPs' user policies.

[2279] Ofcom, 2022. Ofcom's first year of video-sharing platform regulation.

[2280] Predictive search functionalities are algorithmic features embedded in the search bar of a search service.

[2281] Samaritans' response to Ofcom's 2022 Call for Evidence. Samaritans is the UK and Ireland's largest suicide prevention charity.

[2282] The Antisemitism Policy Trust response to Ofcom's 2022 Call for Evidence noted that "*Google's Search autocomplete algorithm has been found to suggest antisemitic, racist and sexist content to users and that Microsoft Bing has been found to direct users to hateful searches via autocomplete*".

found, through their systems, to direct people to hate material and racist content that is legal but can easily direct users to more extreme and illegal content when they follow search prompts."[2283] The latter is also substantiated by an investigation by The Observer in 2016.[2284] There is also research that points to similar risks regarding child sexual abuse material (CSAM).[2285] [2286] A 2019 report by the Antisemitism Policy Trust and Community Security Trust found that Google's removal of a specific antisemitic predictive search suggestion resulted in 10% (one in ten) fewer search requests related to that suggestion in the 12 months following its removal compared to the 12 months prior.[2287] This indicates that the removal of suggestions deemed to present an illegal content risk could materially reduce the likelihood of users encountering illegal content in search results.

Web indexing

25.31 Web indexing[2288] may also present risks of harm to individuals. Through the indexing process, search services act as a gateway to the entire contents of the clear web, including URLs and images that contain illegal content. Evidence indicates that this risk is heightened for CSAM, especially as search engines have been shown to be a common way of finding this type of illegal content.[2289] [2290] Research has found this material could be reached within three clicks on mainstream search engines.[2291] Likewise, content that celebrates, glorifies, or instructs users on self-injurious behaviour (including suicide) has found to have been accessed within a single click from the main search results page.[2292] While not all search results of this nature will amount to the priority offence of encouraging or assisting suicide, the line between legal and illegal in this context is very difficult to draw and we therefore

---

[2283] Antisemitism Policy Trust response to Ofcom's 2022 Call for Evidence.

[2284] Note: The Guardian later reported that Google had altered autocomplete in response and removed some suggestions, while others remained. Source: Cadwalladr, C., 2016. Google, democracy and the truth about internet search, *The Observer*, 4 December. [accessed 4 October 2023]; Gibbs, S., 2016, Google alters search autocomplete to remove 'are Jews evil' suggestion, *The Guardian*, 5 December. [accessed 4 October 2023].

[2285] Note: Microsoft removed the offending suggestions in response. Source: TechCrunch (Constine, J.), 2019. Microsoft Bing not only shows child sexual abuse, it suggests it. [accessed 4 October 2023].

[2286] The WeProtect Global Alliance notes that algorithms that suggest CSAM can have the effect *of "encouraging or inspiring new offending, as well as increasing re-victimisation of those victims of abuse".* WeProtect, 2020. Voluntary Principles to Counter Online Child sexual Exploitation and abuse. [accessed 4 October 2023].

[2287] Antisemitism Policy Trust, Community Security Trust (Stephens-Davidowitz, S.). 2019. Hidden Hate: What Google searches tell us about antisemitism today. [accessed 11 October 2024].

[2288] Indexing is the process of collecting, parsing, and storing data to facilitate fast and accurate information retrieval.

[2289] Steel, C.M.S, 2015. Web-based child pornography: The global impact of deterrence efforts and its consumption on mobile platforms. Child Abuse & Neglect, 44, pp.150-158. [accessed 4 October 2023].

[2290] A qualitative study on the pathways for accessing CSAM online conducted interviews with 20 people who had viewed CSAM online and had been investigated by law enforcement. When asked about their initial exposure, two of the respondents reported that initial exposure occurred through intentional searches on search engines, and when asked about access methods, 13 responded reported using search engines as a pathway to access CSAM. Source: Bailey, A., Allen, L., Stevens, E., Dervley, R., Findlater, D. and Wefers, S., 2022. Pathways and Prevention for Indecent Images of Children Offending: A Qualitative Study. *Sexual Offending: Theory, Research, and Prevention*, *17*, pp.1-24. [accessed 4 October 2023].

[2291] The National Crime Agency carried out research on the availability of CSAM on mainstream search engines and found that access is discoverable within three clicks. UK Government, 2020. Interim code of practice on online child sexual exploitation and abuse. [accessed 4 October 2023].

[2292] The research by the Network Contagion Research Institute (NCRI) found that 22% of the 37,647 individual search results links they assessed across five search engine services contained content that celebrates, glorifies, or instructs self-injurious behaviour within a single click from the main search results page. The report defined self-injurious behaviour as non-suicidal self-injury, suicide, and eating disorders. The research found that 1,580 links were likely to be in scope (promoting self-injury) of extreme (encouraging others to engage in self-injurious behaviour). Source: Network Contagion Research Institute, 2024. One Click Away: A Study on the Prevalence of Non-Suicidal Self Injury, Suicide, and Eating Disorder Content Accessible by Search Engines. [accessed October 2024].

consider it evidence of a clear risk. Further evidence on the risks of encountering illegal content via search services can be found in the Register of Risks chapter 'Search'.

Recommender systems (U2U)

25.32 There are a number of ways by which illegal content can be disseminated on an online service; content recommender systems[2293] are one of these ways. Recommender systems are deployed across many types of U2U service and are often essential to helping users find content they enjoy and wish to engage with. However, where illegal content is uploaded or shared to a U2U service, and missed by any content moderation procedures that are engaged at the point of upload, recommender systems may play a role in amplifying the reach of that illegal content and increasing the number of people who encounter it.[2294] Further information as to how services can mitigate the risks described here can be found in the Codes of Practice (Volume 2).

25.33 The way in which these systems are designed can influence the extent to which illegal content is recommended to users, and that there is a risk of service providers making design adjustments to their systems in a way that results in users being more likely to be exposed to illegal content. A working group review from the Global Internet Forum to Counter Terrorism highlighted that there is a consensus among experts in the technology, government, civil society, and academic sectors that support this claim.[2295] While the focus of these studies tends to be on harmful content, our view is that illegal content – if not caught by moderation processes – would be disseminated in a similar way, given how recommender systems rank and curate content.[2296]

25.34 To back up these claims, there is specific evidence of recommender systems disseminating illegal content, such as CSAM, terrorism and suicide and self-harm content. An investigation by journalists Cook and Murdock (2020) identified that users on one platform could be led on recommendation trails from soft-core pornography to content featuring partially clothed minors. The study found that there was a progression of recommendations from videos showing adult nudity to those featuring minors in sexualised contexts.[2297] Evidence also showed how the recommender system of one social media service had recommended Islamic State accounts to users.[2298] However, it has also been demonstrated that while some recommender systems may disseminate extremist and so-called 'fringe' content, others do not, indicating the importance of different design choices on the risk of encountering harmful and illegal content.[2299] A report by the Molly Rose Foundation

---

[2293] A content recommender system is an algorithmic system which determines the relative ranking of an identified pool of content that includes regulated user-generated content from multiple users on content feeds.
[2294] Ofcom, 2023. Evaluating recommender systems in relation to illegal and harmful content. [accessed 11 November 2024].
[2295] Global Internet Forum to Counter Terrorism, 2021. Content-Sharing Algorithms, Processes, and Positive Interventions Working Group: Part 1. [accessed 27 September 2023].
[2296] Ofcom, 2023. Evaluating recommender systems in relation to illegal and harmful content. [accessed 4 October 2023].
[2297] Cook, J. and Murdock, S., 2020. YouTube Is A Pedophile's Paradise, *HuffPost, 20 March*. [accessed 4 October 2023].
[2298] Waters, G. and Postings, R., 2018. Spiders of the Caliphate: Mapping the Islamic State's Global support network on Facebook. Counter Extremism Project; Ofcom, 2023. Evaluating recommender systems in relation to illegal and harmful content. [accessed 4 October 2023].
[2299] Digital avatar accounts were used to examine how recommender systems affected user exposure to extremist content on different platforms. It found that while some platform recommender systems did disseminate extremist and so-called 'fringe' content, others did not, indicating the importance of different design choices on the risk of encountering harmful.

commented on the speed at which a service's recommender system identified the account's preferences, noting that a large video-sharing platform's For You Page "rapidly identified our interest in suicide and self-harm related material, and we were quickly presented with a range of disturbing and potentially harmful videos." [2300]

Reporting and complaints (U2U and Search)

25.35 Service providers may put in place reporting and/or complaints procedures, including for complaints which are appeals, to mitigate the risks of harm discussed in the Register, and to identify illegal content that has been undetected by moderation systems. However, if these procedures and processes are implemented poorly, they may prove ineffective at facilitating appropriate moderation of illegal content and reducing the risk of harm to users. For example, users may find it difficult to report if the process is too complicated, lengthy or unclear. Further information as to how services can implement reporting and complaints procedures effectively, and mitigate the risks described here, can be found in the Codes of Practice (Volume 2).

25.36 Ofcom's Video-Sharing Platforms Call for Evidence highlighted that services do operate dedicated reporting, such as 'trusted flagger' programmes that enable law enforcers, civil society, charities and other stakeholders to alert services to harmful content,[2301] but these are often applied inconsistently and may fail to address complex kinds of illegal harms like fraud.

Accessibility of reporting/complaints functionalities

25.37 There is a greater risk of illegal content not being appropriately actioned if users struggle to submit complaints, increasing the likelihood of harm to more users who may then come across it or be subject to repeated exposure.

25.38 Children in particular do not report or complain about violative content if the reporting channels are unclear, or if the process is too time-consuming or complicated. Evidence suggests that children are put off complaining or reporting because they don't know how to

Source: Whittaker, J., Looney, S., Reed, A. and Votta, F., 2021. Recommender systems and the amplification of extremist content, Internet Policy Review, 10(2). [accessed 4 October 2023].

[2300] Note: In this study the researchers explored Instagram, TikTok, and Pinterest with avatar accounts registered as being 15-years-of-age. Content was identified and scraped using hashtags that have been frequently used to post suicide and self-harm related material. While this is a singular study and may not represent all children's experiences, it demonstrates that this type of content was available on the services at the time of the study. Source: Molly Rose Foundation, 2023. Preventable yet pervasive: The prevalence and characteristics of harmful content, including suicide and self-harm material, on Instagram, TikTok and Pinterest. [accessed 17 April 2024].

[2301] Tech UK response to 2020 Video-Sharing Platform Regulation Call For Evidence , p.3. [accessed 31 August 2023].

complain, or believe it will be difficult.[2302] [2303] This can hinder content takedown, for instance, in the case of CSEA.[2304]

25.39 In some cases, ineffective reporting and complaints procedures result in service providers being unable to collect reliable information on the type of harmful content and behaviour appearing on their service and are, therefore, unable to take action against further harm.[2305] There is the risk of valid complaints by users not being upheld because users are unable to submit supporting information. Context can help content moderators to make an informed judgement about a complaint and correctly identify illegal content.[2306] The ability to provide additional context can make reporting and complaints systems easier to use by reducing the burden of reporting. In response to our 2022 Illegal Harms Call for Evidence, Refuge highlighted that "survivors must usually report individual pieces of content in turn. Perpetrators will often send dozens or hundreds of messages, making reporting time-consuming and potentially [a] re-traumatising process for survivors".[2307]

Speed of action

25.40 Where users perceive a lack of action on a complaint, or an absence of any action at all,[2308] they are less likely to report in the future. This further increases the risk of harm to individuals.[2309] For example, women who have suffered online abuse can wait months or years for any action to be taken, if it is taken at all, which can increase the stress and

---

[2302] Many children do not know how to use reporting systems or find them difficult to use. Ofcom, 2023, Children's Media Use and Attitudes study: Children's Online Knowledge and Understanding Survey, QC57

[2303] *"Of the 91% of 8-17s who would tell someone if they saw something worrying or nasty online, only 6% would tell the website/app where they encountered the harmful content/behaviour about what they had seen. 35% of 12-17s said they knew how to use a reporting or flagging function, and of those only 14% said they had used it before*" Source: Children and parents: Media Use and Attitudes Report 2022; Of the 88 responses to Ofcom's Call for Evidence: First phase of online safety regulation referring to the need for user reporting to be easy to use for those with vulnerabilities, 14 responses related specifically to children. Some responses highlighted the lack of trust that children have due to problems being difficult to report, or their belief that nothing would be done.

[2304] A report by the Canadian Centre for Child Protection analysed reporting functions across major social media sites and found that CSEA reporting is inaccessible on most, with a lack of specific reporting functions for CSEA, specifically: when reporting nudity; a lack of user (not just content) reporting features; less comprehensive reporting features on mobile compared to desktop apps; and an inability to add contextual information to reports, all deemed to contribute to a greater risk of harm. Source: The Canadian Centre for Child Protection, 2020. Reviewing child sexual abuse material reporting functions on popular platforms: executive summary. [accessed 4 October 2023].

[2305] Online services sometimes report that they struggle to identify activity as foreign interference because of a lack of information to determine whether it is state-linked. The adversarial nature of foreign interference makes attribution a constant challenge. For example, a recent tactic deployed by those engaging in foreign interference is to employ journalists, influencers or 'dark' PR firms to carry out their desired influence activities for them. Without third-party information (e.g. from journalists or information from governments) attribution is difficult. Source: Empirical Studies of Conflict Project (Martin, D. A., Shapiro, J. N., Ilhardt, J.), 2020. Empirical Studies of Conflict Project, Princeton University; Thomas, E., Thompson, N. and Wanless, A. 2020. *The Challenges of Countering Influence Operations*, Partnership for Countering Influence Operations, Policy Perspectives #2, Carnegie Endowment for International Peace. Trends in Online Influence Efforts. [accessed 4 October 2023]; Carnegie Endowment for International Peace (Thomas, E., Thompson, N. and Wanless, A.), 2020. The Challenges of Countering Influence Operations. [accessed 4 October 2023].

[2306] For example, the Integrity Institute highlighted that complaints may be 'denied' without context. Source: Integrity Institute response to May 2024 Consultation on Protecting Children from Harms Online, pp.9-10.

[2307] Refuge response to Ofcom 2022 Call for Evidence: First phase of online safety regulation, pp.7-8.

[2308] Ofcom, 2022. Just one in six young people flag harmful content online. [accessed 25 September 2023].

[2309] Users can often wait a long time to receive any information about their report; e.g., children and women who have suffered online abuse can wait months or years for any action to be taken, if it is taken at all. Children are dissuaded from reporting as the process 'comes to nothing' and they have to chase up reports. Source: Refuge, 2021. Unsocial Spaces. [accessed 4 October 2023].

trauma they may feel.[2310] A study with LGBT+ users found that they fear they will not be taken seriously when reporting content.[2311] Children in particular are often dissuaded from submitting complaints about illegal content as they do not think anything will come of their complaint.[2312]

25.41 On the other hand, benchmarking quick turnarounds for reports arbitrarily may prevent appropriate assessment of illegal content. There is a risk that if providers decide to set themselves arbitrary deadlines to address complaints, they may not assess them accurately. Conversely, providers may also set longer timeframes, which could be perceived as inaction as time progresses and cause frustration to complainants. This has been highlighted as a concern in EU proposals for removal of terrorism content within specific timeframes.[2313]

Service design and user support (U2U)

25.42 Service design[2314] may in some instances facilitate the risk of illegal content being encountered and shared and therefore increase the risks of harm to users on U2U services. Offence-specific risks of harm associated with service design are outlined in different chapters of this Register of Risks. Often, the examples provided relate to how vulnerable users may be recommended content that is increasingly harmful and potentially illegal. Further information on how services can implement service design effectively and mitigate the risks described here can be found in the Codes of Practice (Volume 2).

25.43 Service design might increase the risks of harm where, for instance, service features allow the impersonation of prominent or influential persons. This may lead to offences like fraud and scams, as well as to foreign interference offences, as outlined in the 'Fraud and financial services offences' and 'Foreign interference' chapters of this Register of Risks. Similarly, proscribed terrorist organisations may access services and set up user profiles, which may lead to the spreading of terrorism content. This in turn can increase the risk of harms to individuals encountering illegal content, as pointed out in the Terrorism chapter.

25.44 However, service providers design systems and processes to mitigate the risks mentioned above; if these are implemented poorly, the risks of harm may increase.

[2310] Refuge notes that concerns about long waiting periods for content to be removed, once reported, can compound stress or trauma experienced in some instances, including online abuse. Source: Refuge, 2021. Unsocial Spaces. [accessed 4 October 2023].

[2311] Galop (Hubbard, L.), 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 4 October 2023].

[2312] Ofcom, 2021. Online Experiences Tracker.

[2313] European Parliament, 2021. New rules adopted for quick and smooth removal of terrorist content online. [accessed 4 October 2023]; Tech Against Terrorism, 2021. The Online Regulation series: European Union (update) [accessed 4 October 2023]; European Commission, 2022. Terrorist content online. [accessed 4 October 2023].

[2314] Service design, in its broadest sense, includes the design of all components that shape a user's end-to-end experience with a service. These components can include the business model/decision-making structures, back-end systems and processes, the user interface, and off-platform interventions. In line with the illegal content safety duties, our recommendations in this area will focus on the following categories of measure identified in section 10(4) of the Act*: "design of functionalities, algorithms and other features", "functionalities allowing users to control the content they encounter"* and "*user support measures"* in the context of preventing users from encountering illegal content online.

Blocking, muting and account strikes procedures

25.45 The availability of user controls on services such as blocking and muting functionalities can be important tools for users to protect themselves from harm.[2315] For services which are designed primarily to promote user interaction through direct contact, the relevant harms often take place through such contact, for instance via direct messaging or commenting on content. Blocking functionality can help users protect themselves from this. In cases of cyberstalking, blocking communications from the offending user account can be one of the most effective methods of protection.[2316] Similarly, in cases of abuse, limiting contact with an abusive account by blocking it can help users protect themselves from harmful behaviour.[2317]

25.46 While these functionalities can reduce the risk of harm to users, ineffective procedures around the use of these functionalities could increase the risk of harm. Evidence suggests that risks of harm may increase when account strikes and blocking procedures are ineffective or unclear;[2318] for example, when service providers do not remove or limit violative user profiles such as those reported from users who have received abuse[2319] or user profiles supplying drugs.[2320] Risks of harm may also increase when strike and blocking procedures are used maliciously by users to target certain groups. There are examples of strike and block functions being misused by users looking to persecute other minority users, including transgender and other LGBTQ+ people.[2321]

25.47 For certain kinds of illegal harm, after content take-down, risk may be presented where a service's systems and processes enable the offending user to continue to use the service, or do not seek to prevent (or are ineffective in preventing) a banned user from returning to the service. This has been linked to child sexual exploitation and abuse (CSEA) offences[2322] [2323] and incitement to violence,[2324] but has also been observed as a trend for violative

---

[2315] 66% of respondents to a 2021 study from Thorn reacted to a harmful online experience by blocking the user and 27% muted the user. Source: Thorn, 2021. Responding to Online Threats: Perspectives on Disclosing, Reporting, and Blocking. [accessed 22 October 2024].

[2316] Tokunaga, R. S. and Aune, K. S., 2017. Cyber-Defense: A Taxonomy of Tactics for Managing Cyberstalking. Journal of Interpersonal Violence, 32 (10). [accessed 24 October 2024].

[2317] Pen America, Online harassment Field Manual; *Blocking, Muting and Restricting*. [accessed 24 October 2024].

[2318] Thorn, 2021. Responding to online threats: minors' perspectives on disclosing, reporting and blocking. [accessed 21 September 2023].

[2319] A survey carried out by Refuge found that 15% of female survivors who had experienced harassment and abuse online said: '*the abuse worsened when they reported the perpetrator or took an action to mitigate the abuse, such as blocking the perpetrator online'.* The survey was carried out with 2,264 UK adults, including 1,158 females. 36% of females reported experiencing at least one behaviour suggestive of online abuse or harassment. Source: Unsocial Spaces. [accessed 4 October 2023].

[2320] Volteface found that profiles suspected of supplying drugs on social media sites had multiple back-up accounts in case their current active account was closed. Source: Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2023].

[2321] Business Insider, 2015. Transgender Tinder Users Reported and Banned. [accessed 4 October 2023].

[2322] A Meta report into intent of CSAM sharers *showed "patterns of persistent, conscious engagement with CSAM and other minor-sexualising content if it existed"* when 200 accounts that were reported to NCMEC were analysed. **Sou**rce: Meta, 2021. Understanding the intentions of Child Sexual Abuse Material (CSAM) sharers. [accessed 27 March 2023].

[2323] One research study found that, of a group of 78 perpetrators of child sexual abuse, 42% had attempted to collect all images in an abuse series, or of an individual, indicating a likelihood of persistent offending. Source: Steel, M.S., Newman, E., O'Rourke, S. and Quayle, E., 2021. Collecting and viewing behaviors of child sexual exploitation material offenders — University of Edinburgh Research Explorer. [accessed 4 October 2023].

[2324] This report refers to internal documents from the terrorist group that highlighted the importance of its remaining an influential presence on social media to continue spreading the ISIS message. Source: Berger, J. M. and Morgan, J., 2015. The ISIS Twitter Census: Defining and describing the population of ISIS supporters on Twitter. [accessed 4 October 2023].

behaviour in general.[2325] Similarly, proscribed terrorist organisations with access to services can lead to the spreading of terrorism content. This in turn can increase the risks of harm to individuals encountering illegal content, as pointed out in the Terrorism chapter.

25.48 While transparent systems and processes can increase user understanding of account striking and blocking processes, it is possible that they may increase the risks of harm if they provide perpetrators with better knowledge of how to evade enforcement.

Safety default settings

25.49 Certain functionalities, such as direct messaging, connection lists and the automated display of location information can risk exposure to harm for some users. This is particularly the case for children who can be vulnerable to encountering illegal harm and grooming for the purposes of sexual abuse. Perpetrators of grooming often exploit these functionalities and deploy certain strategies to facilitate the grooming of children, with the intention of sexually abusing them either offline or in person, as explored in our Codes of Practice (Volume 2). Implementing certain safety settings on child user accounts can help mitigate these risks, by disrupting the grooming process and making it more difficult for perpetrators to exploit these functionalities.

25.50 Direct messaging functionalities can be exploited for grooming offences, as perpetrators often develop relationships with children away from public view and parental supervision.[2326] Evidence suggests that in nearly three quarters of cases (74%) where children are contacted online by someone they do not know in person, this contact involves private messaging.[2327] Perpetrators often use direct messaging functionalities to send children unsolicited messages, either for the purposes of committing a grooming offence or to engage in other forms of communication that could increase the risk of harm to a child in relation to other offences.[2328]

25.51 There is evidence across a range of different contexts that suggests that setting defaults is effective at influencing choices and behaviours.[2329] Research indicates that when presented with pre-set courses of action, people are generally more likely to stick with the default option than choose another one.[2330] Our own behavioural research into user content controls also found default settings to be 'sticky' in that, even when prompted, few participants moved away from the content control default setting.[2331]

[2325] In an enforcement update TikTok shared that 90% of repeat violators violate use of the same feature consistently, and over 75% violate the same policy category repeatedly. This demonstrates that continued access for users who commit some kinds of illegal harms poses a high risk, for services of those kinds, of illegal harms being repeated. Source: TikTok (de Bailliencourt, J.), 2023. Supporting creators with an updated account enforcement system. [accessed 27 March 2023].
[2326] See the Grooming chapter of the Register of Risks.
[2327] Office for National Statistics, 2021. Children's online behaviour in England and Wales: year ending 2020. [accessed 15 October 2024].
[2328] This could include, for example, illegal harm related to threats, harassment, stalking, abuse (including hate), coercive and controlling behaviour (CCB), and even terrorism.
[2329] Jachimowicz, J., Duncan, S., Weber, E., and Johnson, E., 2019. When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 3(2), pp.159-186. [accessed 29 November 2024]; Mertens, S., Herberz, M.,Hahnel, U.J.J., and Brosch, T. (2021). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioural domains. *Proceedings of the National Academy of Sciences*, 119(1). [accessed 29 November 2024].
[2330] Thaler, R. H., Sunstein, C. R., and Balz, J. P., 2013, Choice architecture. In E. Shafir (Ed.), The behavioral foundations of public policy (pp. 428-439). Princeton, NJ: Princeton University Press.
[2331] Ofcom, 2024. Behavioural insights to empower social media users. Testing tools to help users control what they see, Behavioural Insights Discussion Paper.

Labelling of user profiles

25.52 Many online services run schemes to label the profiles of notable users (such as celebrities, political figures, media or financial institutions) and say the label can help other users identify the 'authentic' presence or profile of these users on the service.[2332] Some online services have begun to offer monetised schemes whereby users can have their account labelled and get access to special features in return for payment.

25.53 However, confusion about user labelling, or poorly-operated schemes, could cause harm to users. In worst-case scenarios, a labelled account could give a sense of credibility and an amplified voice to a bad actor who is seeking to commit fraud or foreign interference. As a recent example, Martin Lewis, the Executive Chair of the UK's biggest consumer help site, tweeted that an account with a Twitter Blue subscription checkmark was impersonating him to promote a cryptocurrency and stated that 'the blue tick verified fraudulent scammers account' was pretending to be him.[2333] It has also been reported that labelling schemes may be misused to impersonate conspiracy theorists and far-right activists.[2334] [2335] [2336] [2337]

Terms of service and publicly available statements (U2U and Search)

25.54 Reading terms of service and publicly available statements[2338] can help the user understand the rules of use for the service. However, these terms are ineffective if they are not

---

[2332] Verification and labelling schemes refers to schemes operated by services to verify the accounts of certain users, such as notable users or those who subscribe to a paid-for scheme. These schemes may involve labelling a user's profile to indicate that it is verified. Verification in this context may take different forms but usually involves the service carrying out a process before a user profile is labelled as being part of a particular scheme.

[2333] Farrell, L, 2023. Martin Lewis issues warning over fake Twitter account after major change to app. *Daily Record,* 4 April. [accessed 4 October 2023].

[2334] Sardarizadeh, S, 2022. Twitter chaos after wave of blue tick impersonations. BBC News, 12 November. [accessed 4 October 2023].

[2335] Our Media Use and Attitudes trackers look at the experience of UK users online and their attitudes towards this. Participants were asked to judge whether a social media post appeared to be genuine and why they came to their conclusion. This research involved showing social media users a real social media post and asking them if they thought the post was genuine or not, and to give their reasons for doing so. Of the 44% of adult social media users who correctly identified a Money Saving Expert Facebook post as genuine, 51% identified the verification tick as among their reasons for making this judgement. Source: Ofcom, 2023. Adults' Media Use and Attitudes report 2023. [accessed 4 October 2023].

[2336] Respondents aged 12-17 who go online were shown a real NHS Instagram post and asked whether they thought it was genuine or not, and to give their reasons for their opinion. Of the 80% of who correctly recognised that it was a genuine post, nearly three in ten identified the inclusion of a verification tick as one of the factors behind this judgment. Ofcom, 2023. Children and Parents: Media Use and Attitudes. [accessed 4 October 2023].

[2337] Respondents were asked *"when using social media platforms, how often, if at all, do you look out for these kinds of labels (e.g. a tick on a profile) when deciding to follow or interact with an account?*". Nearly three in ten respondents (28%) claimed they 'always' (2%), 'often' (7%) or 'sometimes' (19%) used verification labels when deciding to follow or interact with an account on social media. A further fifth (22%) said they used these labels 'rarely', suggesting that these respondents may find verification labels helpful in certain contexts or situations. Ofcom, 2023. Verification schemes to label accounts poll via YouGov panel. [accessed 4 October 2023].

[2338] On defining terms of service, the Act includes duties that apply in relation to: a) U2U services' terms of service ('terms'), meaning "all documents (whatever they are called) comprising the contract for use of the service (or of part of it) by United Kingdom users" (source: Section 236 of the Online Safety Act 2023); b) search services' publicly available statements ('statements'): search services are required to produce and make available to members of the public in the United Kingdom, a statement setting out certain information about how they operate (source: section 236 of the Online Safety Act 2023); and c) combined services, which have both functionalities, are permitted to set out what would be required in a publicly available statement in terms of service instead (Source: Section 25(2)(a) of the Online Safety Act 2023).

accessible to all users, including children. We consider that a service's terms of service and publicly available statements should be able to be understood by all, so that they can make better-informed choices about what services to use, and how to stay safe online.[2339] It is reasonable to infer that this should reduce users' risk of being exposed to illegal content on a service. Further information as to how services can present their terms of service and publicly available statements effectively can be found in our Codes of Practice (Volume 4).

25.55 To be effective, a service's terms of service and publicly available statements should be easy to read and easy to find.[2340] However, terms of service are often long, confusing and require advanced reading skills to understand, making them unsuitable for many users, especially children.[2341] [2342] Ofcom found that the providers in scope of our video-sharing platform regulation do not use any techniques to improve users' engagement with their terms and conditions of use to help users understand them.[2343] Two-thirds (67%) of UK internet users say that they usually accept terms and conditions without reading them when visiting websites or apps.[2344]

25.56 Clear and accessible terms of service and publicly available statements ensure users can find reliable and up-to-date information about the safety practices of regulated service providers. Certain design choices can help with this. There are a range of techniques which have been shown to be effective at improving user understanding of terms and statements, such as using icons and summaries to aid comprehension.[2345] [2346]

[2339] Several stakeholders mentioned in response to our November 2023 consultation that consideration should be given to ensuring that terms and statements are clear and accessible for children, and users who have disabilities or learning difficulties. Source: Scottish Government response to November 2023 Consultation, p.9. Scottish Government response to May 2024 Consultation, p.17; Parenting Focus response to May 2024 Consultation, p.31. Children's Commissioner for England response to November 2023 Consultation, p.22. Ofcom Advisory Council for Northern Ireland response to November 2023 Consultation, p.9; Mencap response to November 2023 Consultation, p.12; The Cyber Helpline response to November 2023 Consultation, p.16; Glitch response to November 2023 Consultation, p.10. Children's Commissioner for England response to November 2023 Consultation, p.22.

[2340] The ICO's Age Appropriate Design Code states that for terms of service to be accessible to children, they must be prominent, visible and easy to find. Source: ICO, 2020. Age appropriate design: a code of practice for online services. [accessed 17 April 2024].

[2341] Ofcom calculated a 'reading ease' score for the terms of service of the providers in scope of our video sharing platform regulation. All but one was assessed as being "difficult to read and best understood by high-school graduates." Source: Ofcom, 2023. Regulating video sharing platforms (VSPs) - Our first 2023 report: What we've learnt about VSPs' user policies.

[2342] 5Rights for example reported that when they looked at 123 privacy policies for websites likely to be accessed by children, only 9 (7%) had a specific policy targeted at children. Source: 5Rights, 2021. Tick to Agree - Age appropriate presentation of published terms. [accessed 17 April 2024].

[2343] Ofcom, 2023. Regulating video sharing platforms (VSPs) - Our first 2023 report: What we've learnt about VSPs' user policies

[2344] Only 6% of UK internet users aged 16+ said they always read terms and conditions. Source: Ofcom, 2022. Adults' Media Literacy Tracker (table 66). And 33% of UK internet users aged 16-24 reported having ever needed to access social media terms and conditions. Source: Ofcom, 2023. Platform Terms and Accessibility (Q1).

[2345] The Behavioural Insights Team, 2019. Best practice guide: Improving consumer understanding of contractual terms and privacy policies: evidence-based actions for businesses. p.12 [accessed 7 October 2024]. We note that BIT found that using icons with long blocks of text did not work very well. They compared a long privacy policy with no icons to an identical policy that was illustrated with over 20 icons but found that icons did not help customers understand the policy better in that case. This points to the importance of combining icons with short, easy to understand information.

[2346] Danish Competition and Consumer Authority, 2018. Improving the effectiveness of terms and conditions in online trade. *Competitive Markets and Consumer Welfare*, 15, p.5 [accessed 7 October 2024].

25.57 Ensuring accessibility for those with disabilities, and/or those relying on screen-reading technology, is important.[2347] Users with visual or motor impairments may be dependent on using a keyboard to navigate apps and webpages, while screen readers make content on a screen accessible for those who are unable to see it.[2348] [2349] We consider that being able to easily access, and repeatedly visit, terms of service and publicly available statements can help to reinforce users' understanding of their rights and responsibilities as a service user.

User access (U2U)

25.58 Limiting a user's access to a service can mitigate the risks associated to specific accounts that spread illegal content, including by controlling access throughout the user journey. We consider restrictions on user access are a potential means of reducing harm as they can constrain perpetrators from using a service and act as a deterrent against engaging in illegal conduct online. However, there are also risks associated to limiting a user's access, including both risks to freedom of expression and increasing the risk of harm, in cases where a service does not implement its banning process accurately. These require appropriate consideration.

25.59 Some user-access-related functionalities may in some instances increase the risks of harm.[2350] This is outlined in detail in offence-specific chapters of the Register of Risks, such as Drugs and psychoactive substances, Child sexual exploitation and abuse (CSEA) and Harassment, stalking, threats and abuse. Services may put in place systems and processes to mitigate risks of harms linked to a user's access to a service; however, additional risks may arise based on how user access systems and processes are implemented. We outline this is in our analysis below. Further information as to how services can mitigate the risks described here can be found in our Codes of Practice (Volume 2).

25.60 We discuss the risks of blocking and account strikes procedures in the 'Blocking, muting and account strikes procedures' sub-section above.

[2347] Ofcom research found that 18% of internet users aged 16-24 reported having had difficulty reading information online because the content was not keyboard navigable, or was difficult to navigate using a keyboard. The same proportion reported the same difficulty because the content was not compatible, or was difficult to use, with a screen reader or screen-reading technology. Source: Ofcom, 2023. Platform Terms and Accessibility (Q6).
[2348] Web Aim, 2022. Keyboard accessibility. [accessed 7 October 2024].
[2349] Royal National Institute of Blind people, 2023. Screen reading software. [accessed 7 October 2024].
[2350] 'User access' refers to a user's entry into a service and ability to use the functionalities present on that service.

A1. Glossary of terms

This glossary of terms contains definitions for terms used throughout the Register of Risks. These terms may also be referenced in other documents set out for consultation, such as Risk Profiles.

This glossary of terms explains how we have used some key words and phrases in the Register of Risks. It is intended to assist the reader, but to the extent that it simplifies, or is otherwise inconsistent with, any of the legal definitions set out in the Online Safety Act (the "Act"), the definitions in the Act prevail. In case of any conflict between terms used in this glossary and in any Code of Practice, the definition in the Code of Practice takes precedence.

General

Term	Definition
Characteristic	In respect of a regulated service, includes references to its functionalities, user base, business models, governance and other systems and processes.[2351]
Content	Anything communicated by means of an internet service, whether publicly or privately, including written material or messages, oral communications, photographs, videos, visual images, music and data of any description.[2352]
Harm	Means physical or psychological harm. References to harm presented by content, and any other reference to harm in relation to content, have the same meaning given to it by section 235 of the Act.[2353]
Illegal content	Content that amounts to a relevant offence.
Kinds of illegal harms	Refers to harm caused by different categories of relevant offences.
Act	Means the Online Safety Act 2023.
Part 3 or regulated search service	Refers to a search service that falls within the definition of section 4 of the Act.
Part 3 or regulated user-to-user service	A user-to-user service defined in section 4 of the Act.
Priority illegal content	Content that amounts to a priority offence.

[2351] Section 98(11) of the OS Act.
[2352] Section 207(1) of the OS Act.
[2353] Section 201 of the OS Act.

Priority offences	Offences set out in Schedules 5 (Terrorism offences), 6 (CSEA offences) and 7 (Priority offences) to the OS Act.
Relevant offence	Means a priority offence or an offence within the meaning of section 59(4) of the OS Act. This includes both priority offences and non-priority offences.
Risk factor	A characteristic associated with the risk of one or more kinds of harm.
Risk of harm	Means the possibility of individuals encountering harm on a Part 3 service. With reference to a Part 3 U2U service, it means the risk of harm to individuals presented by (a) content on that U2U service that may amount to illegal content; and (b) the use of that U2U service for the commission and/or facilitation of a priority offence. With reference to a Part 3 search service, it refers to the risk of harm to individuals presented by search content on that service that amounts to illegal content.
Search result	In relation to a search service, it means content presented to a user of the service by operation of the search engine in response to a search request made by the user.[2354]
Search services	An internet service that is, or includes, a search engine.
User-to-user services	An internet service by means of which content that is generated directly on the service by a user of the service, or uploaded to or shared on the service by a user of the service, may be encountered by another user, or other users, of the service.

Functionalities and recommender systems

Term	Definition
Accepting cryptocurrency payments	User-to-user service functionality allowing to make and/or receive cryptocurrency payments by means of the service
Accepting online payments	User-to-user service functionality allowing users to make and/or receive financial payments by means of the service
Anonymous user profiles	User-to-user service functionality allowing users to create a user profile where their identity is unknown to an extent. This includes instances

[2354] Section 57(3) of the OS Act.

	where a user's identity[2355] is unknown to other users, for examplebake off through the use of aliases ('pseudonymity'). It also includes where a user's identity may be unknown to a service, for example services that do not require users to register by creating an account.[2356]
Building lists or directories	User-to-user service functionality allowing users to create lists, collections, archives or directories of content or users of the service.
Commenting on content	User-to-user service functionality that allows users to reply to content, or post content in response to another piece of content, visually accessible directly from the original content without navigating away from that content.
Content editing	Functionality type that comprises user-to-user functionalities that allow users to alter user-generated content before or after it is shared.
Content exploring	Functionality type that comprises user-to-user functionalities that allow users to explore and search for user-generated content.
Content recommender systems	Type of recommender system that is used to suggest and curate content that users are likely to find engaging, based on, for example, user preferences and/or history, but also content that is popular and trending on the service at a given moment. Recommender systems exclusively used to suggest goods and services for hire or for sale are a subtype of content recommender system, and we have defined these independently as 'product recommender systems'.
Content storage and capture	Functionality type that comprises user-to-user functionalities that allow users to record and store user-generated content.
Content tagging	User-to-user service functionality allowing users to assign a keyword or term to content that is shared.
Crowdfunding	User-to-user service functionality allowing users to raise money from a large number of users who each contribute a relatively small sum.
Direct messaging	User-to-user service functionality allowing a user to send and receive a message to one recipient at a time and which can only be immediately viewed by that specific recipient.
Downloading content	User-to-user service functionality allowing users to copy content from an online service to their device for local storage.
Editing or deleting posted content	User-to-user service functionality allowing users to modify content that has already been posted the same users or remove it altogether.

[2355] Identity refers to an individual's formal or officially recognised identity.

[2356] The majority of our evidence base speaks of the risks posed by user-to-user anonymity. However, we have indicated where research indicates specifically service-to-user anonymity presents a risk.

Editing usernames	User-to-user service functionality allowing users to alter the name displayed on their user profile.
Editing visual media	User-to-user service functionality that allows users to alter or manipulate images and videos by means of the service.
Encrypted messaging	User-to-user service functionality that allows users to send and receive messages that are end-to-end encrypted.
Ephemeral messaging	User-to-user service functionality that that allows users to send messages that are automatically deleted after they are viewed by the recipient, or after a prescribed period of time has elapsed.
Functionalities	In relation to a user-to-user service, includes any feature that enables interactions of any description between users of the service by means of the service.[2357] In relation to a search service, includes (in particular): (a) a feature that enables users to search websites or databases; (b) a feature that makes suggestions relating to users' search requests (predictive search functionality).[2358] In practice, when referring to functionalities in the Register of Risks, functionalities refer to front-end features of a service. For user-to-user services, functionalities refer to features that enable interaction between users. Functionalities for search services refer to features that enable users to search websites or databases, as well as features that make suggestions relating to users' search requests.
Functionality type	Grouping of functionalities allowing users to engage in a similar online activity.
Group messaging	User-to-user service functionality allowing users to send and receive messages through a closed channel of communication to more than one recipient at a time.
Hyperlinking	User-to-user service functionality enabling users to access other internet services by clicking or tapping on content present on the service.
Live audio	User-to-user service functionality that allows users to communicate with one another in real-time through speech or other sounds.[2359]

[2357] Section 233(1) of the OS Act. Please refer to section 233(2) of the OS Act for a non-comprehensive list of user-to-user functionalities.
[2358] Section 233(3) of the OS Act.
[2359] While one-to-one live aural communications are not regulated-user generated content, they may be in scope of the OS Act when 'accompanied by user generated content of any other kind, except identifying content' or when they are recorded. Aural communications can be regulated user-generated content if they allow more than two users to communicate by means of the service. Section (55)5 of the OS Act.

Livestreaming	User-to-user service functionality that allows users to simultaneously create and broadcast online streaming media in, or very close to, real time.
Network recommender systems	Type of recommender system that suggests users and/or groups of other users to connect with. Network recommenders may consider a variety of user interactions, mutual connections, and group memberships to determine which network recommendations might be relevant and useful.
Posting content	User-to-user service functionality allowing users to upload and share content on open channels of communication.
Posting goods or services for sale	User-to-user service functionality allowing users to post content dedicated to offering goods and services for sale. This does not include paid-for advertisements,[2360] but may serve the function of allowing users to promote goods or services
Posting or sending location information	User-to-user service functionality allowing users to share their current or historic location, record a user's movement, or identify which other users of the service are nearby.
Product recommender systems	Under the Act's definition of user generated content, product recommender systems are considered a subtype of content recommender systems. This is because product listings on U2U online marketplace and listing services are considered user-generated content. Product recommender systems are exclusively used to suggest good and services for sale or for hire. Product recommender systems suggest goods and services that a user might want to purchase. These recommendations are typically based on search and purchasing history.
Reacting to content	User-to-user service functionality allowing users to express a reaction, such as approval or disapproval, of content that is shared by other users through dedicated features that can be clicked or tapped by users.[2361]
Recommender systems	An algorithmic system which, by means of a machine learning model, determine the relative ranking of suggestions made to users on a U2U service. The overarching objective of recommender systems is to ensure users receive suggestions they are likely to find relevant and engaging, thereby improving allocative efficiency in the digital marketplace. This can include suggesting connections, groups, events, and content.
Re-posting or forwarding content	User-to-user service functionality that allows users to re-share content that has already been shared by a user.

[2360] See 'advertising-based revenue model' in business models for more information.
[2361] This for instance includes 'liking' or 'disliking' a post.

Reverse image searching	Search service functionality enabling users to find similar images based on sample images used as a search query.
Screen capturing or recording	User-to-user service functionality that allows users to capture an image or record a video showing the contents of their display.[2362]
Search prediction and personalisation	Functionality type that comprises search service functionalities allowing suggestions to be made relating to users' search requests.
Search query inputs	Search service functionality type by means of which users input search queries.
Transactions and offers	Functionality type that comprises user-to-user service functionalities that allow users to buy, sell, and exchange goods and services with each other. Includes non-profit transactions and offers.
User communication	Functionality type that comprises user-to-user service functionalities that allow users to communicate with one another either synchronously or asynchronously. Includes communication across open and closed[2363] channels.
User connections	User-to-user service functionality that allows users to follow or subscribe to other users. Users must sometimes be connected to view all or some of the content that each user shares.
User events	User-to-user service functionality that enables users to create an online space to share content that is dedicated to a particular event. This can include a date, description, and attendance of users.
User generated content search filtering	User-to-user service functionality allowing users to narrow the parameters of the returned search results, which display user-generated content.
User generated content searching	User-to-user service functionality allowing users to search for user generated content by means of a user-to-user service
User groups	User-to-user service functionality allowing users to create online spaces that are often devoted to sharing content surrounding a particular topic. User groups are generally closed to the public and require an invitation or approval from existing members to gain access. However, in some cases they may be open to the public.
User identification	Functionality type that comprises user-to-user service functionalities that allow users can identify themselves to other users.

[2362] While users can often record or capture content using third-party services, screen recordings and captures are often shared on user-to-user services as user-generated content and some user-to-user services have dedicated screen recording and screen capturing functionalities.

[2363] See content audiences for definition of open and closed channels of communication.

User networking	Functionality type that comprises user-to-user service functionalities that allow users to find or encounter each other and establish contact.
User profiles	User-to-user service functionality that is associated with a user account, that represents a collection of information shared by a user which may be viewed by other users of the service. This can include information such as username, biography, profile picture, etc., as well as user-generated content generated, shared or uploaded by the user using the relevant account. [2364][2365]
User searching	User-to-user service functionality that enables users to search for other users of a service.
User tagging	User-to-user service functionality allowing users to assign other users, typically by their username, to content that is shared.
Video calling	User-to-user service functionality allowing users to communicate with one another in real-time through video communications.

Business models and commercial profile

Term	Definition
Advertising-based revenue models	Revenue models that generate income through payments for the display of advertisements promoting a product or service.
Business models	Way in which a business operates to achieve its goals. For the purposes of this risk assessment, this includes a service's revenue model and growth strategy. [2366]
Commercial profile	Size of the service in terms of capacity,[2367] the stage of service maturity and rate of growth in relation to users or revenue
Early-stage services	Services in the initial phases of their lifecycle, typically encompassing the startup and early growth stages. This is characterized by its early establishment, limited operational history, and ongoing efforts to establish itself in the market
Growth strategy	How the service plans to expand its business. For example, through growing revenue and number of users.

[2365] Users can sometimes create fake user profiles, which are not a functionality in themselves, but are user profiles that impersonates another entity or are intentionally misleading.
[2365] Users can sometimes create fake user profiles, which are not a functionality in themselves, but are user profiles that impersonates another entity or are intentionally misleading.
[2366] 'Business model' can be defined more widely to describe the way in which a service creates value to its users (value proposition), how it delivers this value to users, and how it captures value for itself. However, we adopt a narrow definition in the risk assessment to avoid overlap with the other risk characteristics. This does not affect the overall risk assessment as risk factors that would have been identified under the broader definition are captured elsewhere.
[2367] In terms of number of employees and/or revenue.

High-capacity services	Services with a large number of employees and/or revenue[2368]
Low-capacity services	Services with a small number of employees and/or revenue[2369]
Revenue model	How a service generates income or revenue
Subscription-based revenue models	Revenue models that generate income by selling access (or premium access) to a service for a period of time in return for a fee

User base

Term	Definition
Child user	A user under the age of 18
Protected (user) characteristics	Means age; disability; gender reassignment; marriage and civil partnership; pregnancy and maternity; race; religion or belief; sex; and sexual orientation.[2370]
User base	Users of a service. A user does not need to be registered with a service to be considered a user of that service.[2371]
User base demographics	Demographic make-up of the user base, including selected characteristics, intersectional dynamics and other relevant demographic factors.

Governance, systems and processes

Term	Definition
Account blocking	Process of removing users and often also preventing them from using a service. It is usually deployed for serious or multiple infringements of service policies as it has a high level of impact on the user. A user can be temporarily blocked or have their account and access permanently suspended (often referred to as a 'ban').
Account strikes	Process of adding a mark on a user or their account to note that they have contravened the service's policies.
Content moderation	When a service reviews content to decide whether it is permitted on its platform.
Governance	Structures that ensure the adequate oversight, accountability, and transparency of decisions within a service which impact user safety.

[2368] Our evidence does not currently allow for quantitative thresholds to be drawn for service capacity. Services should nevertheless consider the number of employees and revenue as a risk factor.
[2369] Our evidence does not currently allow for quantitative thresholds to be drawn for service capacity. Services should nevertheless consider the number of employees and revenue as a risk factor.
[2370] Section 4 of the Equality Act 2010.
[2371] Section 195 of the OS Act makes clear that 'it does not matter whether a person is registered to use a service' for them to be considered a 'user.'

	This is in relation to organisational structure as well as product and content governance.
Service design	Design of all components that shape a user's end to end experience with a service. These components can include the business model or decision-making structures, back-end systems and processes, the user interface, and off-platform interventions.
Systems and processes	Characteristic concerning the actions taken by a service, including procedures to mitigate the risk of harm arising from illegal content being encountered. This can be either human or automated, or a combination of the two, and include technology.
User access	A user's entry into a service and ability to use the functionalities present on that service.
Verification and labelling schemes	Schemes operated by services to grant verified status to the profiles of certain users, such as notable users or those who subscribe to a paid-for scheme. These schemes may involve labelling a user's profile to indicate that it is verified. Verification in this context may take different forms but usually involves the service carrying out a process before a user profile is labelled as being part of a particular scheme.

Service type

Term	Definition
Discussion forums and chat room services	A user-to-user service type describing general services that generally allow users to send or post messages that can be read by the public or an open group of people.
Downstream general search service	Search service type describing a subsection of general search services. Downstream general search services provide access to content from across the web, but they are distinct in that they obtain or supplement their search index from other general search services.
File-storage and file-sharing services	User-to-user service type describing services whose primary functionalities involve enabling users to store digital content and share access to that content through links.
Fundraising services	User-to-user service type describing services that typically enable users to create fundraising campaigns and collect donations from users.
General search services	Search service type describing services that enables users to search the internet and which derives search results from an underlying search index (developed by either the service or a third party).
Information-sharing services	User-to-user service type describing services that are primarily focused on providing user-generated informational resources to other users.
Messaging services	A user-to-user service type describing services that are typically centred around the sending and receiving of messages that can only be viewed or read by a specific recipient or group of people.

User-to-user pornography services	User-to-user service type whose principal purpose is to disseminate user-generated pornography.
Dating services	User-to-user service type describing services that enable users to find and communicate with romantic or sexual partners.
Gaming services	User-to-user service type describing services that allow users to interact within partially or fully simulated virtual environments.
Marketplaces and listings services	User-to-user service type describing services that allow users to buy and sell their goods or services.
Payment services	User-to-user service type describing websites or applications that financial payment providers often have that enable users to send and receive money.
Service type	Service type is a characteristic that in general refers to the nature of the service.[2372] This, for instance, includes social media services and private messaging services
Social media services	User-to-user service type describing services that connect users and enable them to build communities around common interests or connections.
Vertical search services	Search service type describing services that enable users to search for specific topics, or products or services offered by third party providers. Unlike general search services, they do not return search results based on an underlying search index. Rather, they use an API or equivalent technical means to directly query selected websites or databases with which they have a contract, and to return search results to users
Video-sharing services	User-to-user service type describing services that allow users to upload and share videos with the public.

Other terms

Term	Definition
Adult Services Website	Marketplace type services which allow for the listing of a sexual service.
Augmented reality	Involves overlaying digital content, which could include a combination of sound, video, text, and graphics, onto a real-world environment using a headset or a device with a camera, such as a mobile phone.

[2372] Certain service types have been selected because our evidence suggests that they can be used to facilitate or commit relevant offences.

Blockchain	A decentralised, distributed ledger that stores the record of ownership of digital assets.[2373]
Bot	An umbrella term that refers to a software application or automated tool that has been programmed by a person to carry out a specific or predefined task without any human intervention.
Clear web	Publicly accessible websites that are indexed by search engines.
Click farming	Practice of manually clicking on online adverts to increase the clickthrough rate value, boost engagement metrics, and inflate impressions. This activity can be carried out by bot accounts, as discussed here, or by large groups of workers.
Content audience	Refers to whether content is shared on open or closed channels of communication. Open channels are areas of services where content is visible to the general public or any user. Closed channels are areas of a service where content is limited to a smaller audience, and where users can expect more privacy, such as direct messaging or user groups that have controls or restrictions on who can join.
Content format	Refers to the format in which content is made available. This, for instance, includes content in the form of images, video, audio, text and emojis.
CSAM URL	A URL at which CSAM is present, or which includes a domain which is entirely or predominantly dedicated to CSAM.
Deepfake	Specific type of media that involves the use of AI algorithms, particularly generative AI models, to modify videos, images or audio to create realistic synthetic content. This is often done by superimposing the face of a person onto the body of another person in a video or image as well as voice manipulation with lip syncing. Deepfakes are commonly shared as user generated content on user-to-user services but could also potentially be created using functionalities present on user-to-user services. Deepfake technology is currently used to create content that can be harmful; however, we acknowledge that it may also have positive use cases.
Down-blousing	Refers to someone taking a photo down a woman's top without consent.[2374]
Generative artificial intelligence	Also known as 'GenAI,' generative artificial intelligence is an emerging form of AI that refers to machine learning models which can create new content in response to a user prompt. These tools can be used to produce text, images, audio, video and code, which closely resemble the broad datasets on which the models are trained.

[2373] Builtin (Daley, S.), 2022. What Is Blockchain? [accessed 18 June 2023].
[2374] Ministry of Justice, 2022. New laws to better protect victims from abuse of intimate images. [accessed 3 August 2023].

Hashtag hijacking	Use of a hashtag for a purpose other than it was created – such as tagging a message containing undesirable or harmful content with a popular, but unrelated, hashtag to surface this content to a target audience.
Indexing	Process of collecting, parsing, and storing of data by a search engine to facilitate fast and accurate information retrieval.
Like farming	Use of fake pages on social media services designed to artificially increase the popularity of a page, so it can be sold to buyers seeing accounts with large followings or for scam and fraud activity.
Mixed reality	Refers to the blending of physical and virtual worlds to produce new environments where physical and digital objects co-exist and interact in real time.
Money mule	Someone who receives money from a third party in their bank account and transfers it somewhere else, or who withdraws it as cash and gives it to someone else, obtaining a commission for it or payments in kind. These individuals are targeted by 'money mule recruiters', sometimes referred to as 'mule herders', who recruit money mules.
Financially motivated sexual extortion ('sextortion')	Form of blackmail that involves threatening to publish sexual information, photos or videos about someone. This may be to extort money or to force the victim to do something against their will. Photos or recordings are often made without the victim realising or consenting.[2375]
Trend jacking	Refers to when influencers, brands or organisations insert themselves into conversations online that are gaining a lot of attention – for example, by using associated hashtags or trending audios.
Up-skirting	Refers to someone taking a photograph that appears to have been taken up a person's clothing (such as a skirt) without consent.
Virtual reality	Involves the use of a head mounted display to access a virtual experience, which could be digitally created or a captured 360° photo or video.

[2375] Metropolitan Police, n.d. Sextortion. [accessed 4 August 2023].

A2. Updating the Register of Risks

Introduction

A2.1 Ofcom is required under the Act to conduct a sector-wide risk assessment into the risk of illegal harm to people in the UK. Our findings are set out in the **Illegal Harms Register of Risks ('Register of Risks'),** which we are publishing alongside this statement. Furthermore, ensuring service providers undertake a high quality risk assessment is one of our strategic objectives (see the Introduction to the Regsiter of Risks). We expect service providers to refer to the Register of Risks when they conduct their own risk assessments.

A2.2 The Register of Risks is underpinned by an **extensive, robust and reliable evidence base** that has been used to assess the chararcteristics of services that can pose an increased risk of harm to individuals. For user-to-user (U2U) services, this includes the risk of users encountering illegal content by means of the service, services being used for the facilitation and commission of priority offences and the risk of, as well as the risk of harm to individuals arising from these things. For search services, this includes only the risk of harm from users encountering illegal content.

A2.3 In our November 2023 Illegal Harms Consultation and our August 2024 Further Consultation on Torture and Animal Cruelty ('August 2024 Further Consultation'), our assessment focused on over **130 priority offences** grouped into **15 kinds of illegal harm.** Based on stakeholder responses, our final Register of Risks now contains an expanded and more detailed evidence base of the causes and impacts of harm. Furthermore, we now group the priority offences into **17 kinds of illegal harm**. These include illegal harms such as child sexual exploitation and abuse (CSEA), terrorism, fraud, hate speech, weapons and drugs offences, modern slavery and human trafficking, foreign interference and most recently, animal cruelty. The complete Register of Risks covers the priority and non-priority offences in the Act.

A2.4 The Act also requires Ofcom to prepare and publish **'Risk Profiles'** based on the findings in our Register of Risks.[2376] They are a tool used to summarise our evidence base and highlight important factors relevant to services that may increase the likelihood of harm from illegal content.

A2.5 Service providers are required to **take account of our Risk Profiles when they conduct their own risk assessments**. The Act gives Ofcom discretion about how to do this, including how we group services together.

A2.6 In our November 2023 Consultation, we proposed an approach to Risk Profiles that sought to highlight what **characteristics** of online services are likely to increase risk ('**risk factors**') and indicate which **kinds of illegal harms** would be likely to occur on services with those characteristics.[2377] The Risk Profiles did not set out all the risk factors from the Register of

[2376] Section 98(5) of the Act.

[2377] Characteristics include a service's user base, business model, functionalities and any other matters we deem relevant to risk. Risk Profiles focus predominately on user base demographics, functionalities and business models. Step 2 of the risk assessment guidance provides information for services on user base size, governance, and systems and processes.

Risks, but only those we considered to be particularly important for service providers to consider.

A2.7 We presented these as two distinct risk profiles in table form; one for U2U services and another for Search services. We proposed that services should consult the relevant table and take account of all **general risk factors** and decide which **specific risk factors** are relevant to them. The Risk Profiles should be used as a starting point for service providers to identify risk factors in Step 1 of our Risk Assessment Guidance.

A2.8 As a result of stakeholder feedback, and in response to updated evidence within the Register of Risks, we have subsequently updated the Risk Profiles – in large part **clarifying exactly how they are to be used** and adding **new associations between kinds of illegal harm and certain risk factors**.

A2.9 Below, we explain our proposed methodology for creating the Register of Risks and the Risk Profiles in our November 2023 Consultation proposals and our August 2024 Further Consultation on Torture and Animal Cruelty. We then summarise the key themes of stakeholder responses recevied in realtion to these two products and our decisions.

Methodology

Evidence base

A2.10 Our risk assessment process has consisted of identifying and analysing a significant amount of high-quality evidence from over a thousand individual sources. This has included considering evidence provided in responses to our July 2022 call for evidence, our November 2023 Illegal Harms Consultation and August 2024 Further Consultation which focused specifically on animal cruelty and human torture content. We also looked at relevant Ofcom research, academic papers from a range of disciplines, government bodies including law enforcement, third-party sources, and information from charities and other civil society organisations. Given the wide range of sources we relied on, we have taken steps to quality assure all evidence so that it is robust and reliable according to the following criteria: method, robustness, ethics, independence and narrative.[2378]

A2.11 Our published Register of Risks now includes well over 1000 individual sources, with hundreds more reviewed but unused in the process of creating our assessment of the risk of harm online. Stakeholders provided additional evidence to support this analysis for each of the 17 kinds of priority illegal harm and 6 non-priority communications offences which each have a chapter in the Register of Risks. Where evidence is limited, we have used our judgement and expertise about specific harms to draw conclusions where we think this can help service providers to identify potential risks. We have also referred to evidence related to content or activity that is broader than the offences discussed, where we consider that

[2378] 'Method' examined the strengths and weaknesses of the methodology for that particular topic, such as whether appropriate data collection methods were used. 'Robustness' considered both the size and coverage of the sample, and quality of analysis – for example, how missing data values were accounted for. 'Ethics' refers to how well ethical considerations were addressed in the study, such as how personal data was handled. 'Independence' examined the origins of the research and whether any stakeholder interests might have influenced findings. 'Narrative' refers to the commentary within the report and whether conclusions are sufficiently backed by the research, and whether there is a clear distinction between the findings and the interpretation.

this is still likely to be relevant to those offences, including the risk of a service being used to commit or facilitate a priority offence.

A2.12 Despite the extensive review of evidence, there remained gaps for some relevant offences in our draft Register of Risks which we published in the November 2023 Consultation such as on the links between the characteristics the Act requires us to assess and the kinds of illegal harms. The evidence base has now been expanded in light of consultation responses and is reflected in the final Register of Risks, but the breadth and depth of evidence remains varied across the kinds of illegal harms.

A2.13 As stated in the November 2023 Consultation, for kinds of illegal harm where we have less evidence, we do not take this as an indication that the kind of illegal content in question does not cause harm online, as there may be other factors that contribute to the lack of robust and reliable evidence at this time. In some cases, we have been able to build on our understanding of these illegal harms by engaging with law enforcement, specialist agencies, civil society organisations and expert groups.

A2.14 Lastly, it is worth highlighting two important observations regarding the evidence around 'types' of service. First, we have referred to some evidence that relates to specific services. This is not intended to be a judgement about the trust and safety systems of those services. Rather, we have included such service specific evidence because of what it tells us about certain functionalities and their associated risks generally. Second, we do not have specific evidence relating to all types of U2U services. Where we are lacking this evidence, we have made reasonable inferences about the risks that may arise on them.

Characteristics and risk

A2.15 The Act requires Ofcom to take into account how the characteristics of a service may give rise to risk. The Act defines 'characteristics' broadly as including a service's **functionalities, user base, business model, governance and other systems and processes.** We consider these characteristics both individually and, where relevant, in combination.

A2.16 Most of the characteristics referenced in the Act are not specifically defined. We recognise that given the diversity and range of services in scope of the Act, many services are likely to define some of these concepts differently. In our November 2023 Consultation, we set out definitions we used to conduct our sector-wide risk assessment in Table 1 below.

A2.17 The list of characteristics in the Act is not exhaustive, so it is open to Ofcom to identify other relevant characteristics. We consider our evidence justified including three additional service characteristics that can give rise to risk: **service type, recommender systems** and **commercial profiles**. Definitions of these are also included in Table 1. Our rationale for these additional service characteristics in our November 2023 Consultation was as follows:

- There is some evidence to suggest that certain **service types** with common features and functionalities, are more likely to be used to commit and facilitate some offences.[2379] We have therefore identified some service types as a driver of risk.

[2379] For example, social media services are relevant to most types of illegal harm due to the wide range of functionalities they provide for sharing content and connecting users. In contrast, marketplaces and listing services are associated with fewer kinds of illegal harm, primarily those where the supply or sale of illegal goods or services is particularly important.

- We have also identified **recommender systems** as a relevant characteristic because of the key role they play in determining what content users see and engage with, therefore contributing significantly to a user's experience of services that use them. Recommender systems can be used in many ways which can influence how a user might experience risk of harm on a service. Most commonly this includes content recommender systems designed for the curation of content feeds, and network recommender systems that are used to recommend other users to follow/befriend.
- We have also included **commercial profiles** as our evidence showed that services with certain commercial profiles are likely to have weaker risk management, which can make them targets for perpetrators.

A2.18 There are many characteristics within the categories outlined above. To conduct our risk assessment, we identified a set of characteristics which were considered potential 'risk factors'.

A2.19 We recognise that not all characteristics are inherently harmful; we therefore use the term **'risk factor'** to describe a characteristic for which there is evidence of a risk of harm to individuals. For example, a functionality like livestreaming is not inherently harmful but evidence has shown that it can be abused by perpetrators; when considering specific offences such as terrorism or CSEA, a functionality like livestreaming can give rise to risk of harm or the commission or facilitation of an offence.

A2.20 An important part of our assessment of the causes and impacts of harm focuses on the extent to which individual risk factors play a role in the risk of each kind of illegal harm occurring on online services.

A2.21 Alongside the wealth of written evidence reviewed, we also engaged extensively with external stakeholders including law enforcement, specialist agencies, civil society organisations and expert groups focusing on relevant illegal harms online to ensure we represent harms accurately and correctly attribute risk to certain characteristics.

Table 1. Definitions for each characteristic

Characteristic	Description
Service type	Service types in general refers to the nature of the service. This, for instance, includes social media services, private messaging services, adult services, video sharing services or online gaming services.
User base	The Ofcom risk assessment has considered the size of a service's user base and user base demographics. It includes consideration of both registered and non-registered users of a service.[2380]

[2380] The Act makes clear that 'it does not matter whether a person is registered to use a service' for them to be considered a 'user' (section 227 of the Online Safety Act). The Act is only concerned with the number of 'United Kingdom users' of the service, so where the user is an individual, they count as a user only where they are in the United Kingdom; similarly, where the user is an entity, they count only where they have been formed or incorporated in the United Kingdom (section 227(1) of the Online Safety Act 2023).

Characteristic	Description
Functionalities	An umbrella term for features that are available to users of a service. The Act defines U2U service functionalities as features that enable interaction between users. Functionalities for search services are defined as features that enable users to search websites or databases, as well as features that make suggestions relating to users' search requests. The Act[2381] includes a non-exhaustive list of functionalities. The Ofcom risk assessment has also considered a number of other relevant functionalities in addition to those listed in the Act in our analysis.
Recommender systems	Recommender systems are a type of information retrieval and ranking systems that are designed to personalise and optimise a user's experience of the service to ensure that they are suggested content that they will find engaging. We considered two types: content recommender systems (curates personalised content feeds for users) and network recommender systems (recommends other users and/or groups to follow and connect with).
Business models	Business model, in a broader sense, outlines the way a business operates to achieve its goals. For the purpose of Ofcom's risk assessment, we have adopted a narrow definition that considered two things:[2382] • Revenue models: how the service generates income or revenue e.g. through advertising, subscription, donation, transaction fees etc. • Growth strategy: how the service plans to expand its business. For instance, through growing revenue and number of users.
Commercial profile	We use commercial profile to refer to the size of the service in terms capacity (i.e. revenue and/or number of employees), the stage of service maturity[2383] and rate of growth in relation to users or revenue.

[2381] Section 233: For U2U: (a) creating a user profile, including an anonymous or pseudonymous profile; (b) searching within the service for user-generated content or other users of the service; (c) forwarding content to, or sharing content with, other users of the service; (d) sharing content on other internet services; (e) sending direct messages to or speaking to other users of the service, or interacting with them in another way (for example by playing a game); (f) expressing a view on content, including, for example, by— (i) applying a 'like' or 'dislike' button or other button of that nature, (ii) applying an emoji or symbol of any kind, (iii) engaging in yes/no voting, or (iv) rating or scoring content in any way (including giving star or numerical ratings); (g) sharing current or historic location information with other users of the service, recording a user's movements, or identifying which other users of the service are nearby; (h) following or subscribing to particular kinds of content or particular users of the service; (i) creating lists, collections, archives or directories of content or users of the service; (j) tagging or labelling content present on the service; (k) uploading content relating to goods or services; (l) applying or changing settings on the service which affect the presentation of user-generated content on the service; (m) accessing other internet services through content present on the service (for example through hyperlinks).
For Search: (a) a feature that enables users to search websites or databases; (b) a feature that makes suggestions relating to users' search requests (predictive search functionality).

[2382] 'Business model' can be defined more widely to describe the way in which a service creates value to its users (value proposition), how it delivers this value to users, and how it captures value for itself. However, we adopt a narrow definition in the risk assessment to avoid overlap with the other risk characteristics. This does not affect the overall risk assessment as risk factors that would have been identified under the broader definition are captured elsewhere.

[2383] 'Maturity' refers to the stage the service or company is at in the typical business lifecycle. The stages can be split into four: i) introductory or start-up stage, ii) growth stage, iii) maturity stage, and iv) decline. The maturity stage is characterised by high revenues, cashflow and profitability.

Characteristic	Description
Governance, systems and processes (GSP)	Governance, systems and processes (GSP) are typically put in place to prevent and/or reduce risk; we review how inadequate or absent GSP in a service can lead to risk. We define the terms as follows: **Governance**: Any structure, or structures to ensure that decisions are made with adequate oversight, accountability, transparency and regard to online safety compliance, specifically in relation to risk management, product and content governance within a service. **Systems and processes**: Series of actions taken by a service, including actions that mitigate the risk of harm arising from illegal content being encountered. These may include any human or automated systems or processes, other technologies.

Kinds of illegal harm

A2.22 The relevant offences considered in our Register of Risks are:

- **Priority offences,** which include terrorism offences, offences related to CSEA and other priority offences set out in the Act. So-called '**inchoate offences**'[2384] are also treated as priority offences.
- **Relevant non-priority offences**[2385] including the **Communications offences** (Part 10): false communications offence, threatening communications offence, offences of sending or showing flashing images electronically ('epilepsy trolling'), offence of sending etc photograph or film of genitals ('cyberflashing') and the offence of encouraging or assisting serious self-harm.

A2.23 In the Register of Risks, we group over 130 priority offences referenced in the Act to ensure our detailed analysis is as accessible as possible. This helps us bring out risks of harm from illegal content and activity based on relationships and similarities between the offences as well as reflecting the real lived experiences of people whom services are under a duty to protect.

A2.24 How we group these priority offences into 'kinds of illegal harm' is relevant to all our regulatory products. We group evidence for assessing the risks of harm in the Register of Risks and Risk Profiles on this basis, and thus expect service providers to take account of this when conducting their risk assessments.[2386] The scope of certain recommended meausures in our Codes of Practice will depend on the level of risk of particular kinds of illegal harms on a service. For further details on the relationship between the kinds of

[2384] As explained in Overview of Illegal Harms, 'inchoate offences' include assisting someone else to commit a priority offence, encouraging someone else to commit a priority offence, attempting to commit a priority offence or conspiring to commit a priority offence.

[2385] Referred to in the Act as 'other offences', they are all offences under UK law that are not priority offences, where (a) the victim or intended victim of the offence is an individual (or individuals); (b) the offence is created as a result of the Act, another Act, an order of Council or other relevant instruments; (c) the offence does *not* concern the infringement of intellectual property rights, the safety or quality of goods, or the performance of a service by a person not qualified to perform it; and (d) the offence is *not* an offence under the Consumer Protection from Unfair Trading Regulations 2008.

[2386] However, within each grouping we sometimes refer to individual offences where appropriate, for example where the particular observation or evidence is relevant only to specific offences.

illegal harms in the Register of Risks and the Codes of Practice please see Volume 2 – Service design and user choice.

A2.25 In our November 2023 Consultation, we grouped over 130 priority offences into 15 kinds of illegal harms. We also collated and analysed evidence on a number of non-priority offences with their own chapters in the Rigister of Risks to ensure our analysis covered all offences within the Act.

A2.26 In our August 2024 Further Consultation which focused specifically on animal cruelty and human torture content, we proposed adding two new chapters to the Register of Risks. One on the priority offence relating to animal welfare, which had been added to the Act late in the legislative process such that we could not include it in our November 2023 Consultation – the 'Animal cruelty' chapter. We also considered that it was appropriate to consider a relevant non-priority offence related to additional kinds of content that could be deemed to be illegal human and animal torture content – the 'Obscene content showing torture of humans and animals (the s.127(1) offence)' chapter.

Stakeholder responses and decisions by theme

A2.27 Feedback provided in response to the Register of Risks was extensive, with a wide range of recommendations and supporting evidence provided regarding the links we should draw between certain kinds of illegal harm and the characteristics of online services. As well as providing new and previously uncited evidence, stakeholders highlighted in detail where they felt our analysis and conclusions did not align with their own evidence, experience or views.

A2.28 In response, we have made hundreds of additions and edits to the Register of Risks to incorporate new evidence, update our conclusions in response to this evidence, and to clarify numerous points. This includes filling gaps in our analysis in relation to specific kinds of illegal harm and providing clearer explanations on broad themes that are relevant across illegal harms, such as the role of business models and generative AI. We have summarised these chapter-by-chapter updates in the 'Further stakeholder responses' Annex to Statement: Protecting people from illegal harms online, where we have also set out the small number of instances where we have not made a change despite requests to do so from stakeholders.

A2.29 Below we highlight some themes that emerged from the responses with notable examples of stakeholder responses followed by our decisions.

Support for our methodology and overall approach

A2.30 We did not receive feedback suggesting that our methodology and overall approach to conducting our risk assessment and producing the Register of Risks was incorrect. Although, as noted, stakeholders provided extensive feedback about specific aspects of the Register of Risks and the conclusions drawn from our analysis (discussed below).

A2.31 Stakeholders were in agreement with the overarching need for Ofcom to produce the Register of Risks and the way it was set out, while also providing a range of recommendations and considerations we have addressed. For example:

- Clean up the Internet said *"The Register of Risks provides an excellent assessment of the causes and impacts of online harms, and a good basis for a first set of codes."* But highlighted a need for ongoing updates to the evidence base; *"We would expect it to require regular updating as technology changes and evidence emerges, including as Ofcom exercises its information gathering powers under the Act, which we would encourage Ofcom to do soon and regularly."*
- The Children's Commissioner *"broadly supports the assessment of causes and impacts of online harms contained in the register of risks."* But flagged additional areas to be addressed, such as the inclusion of user age as a characteristic in the assessment of harm relating to terrorism content.
- Meta and WhatsApp said: *"We are grateful for Ofcom's extensive analysis and willingness to engage with providers of services on the cause and impacts of online harms and we acknowledge Ofcom's considerable effort to develop a register of risks based on three years of dedicated work."*
- The Metropolitan Police Service and Counter Terrorism Policing said: *"We note that you articulate a significant evidence base for your findings which correctly reflects our analysis of the online space."*
- The Global Network Initiative said: "*As a general matter, GNI agrees with Ofcom's assessment that the characteristics of a particular service, as well as its governance, systems, and processes for identifying and addressing risks, contribute significantly to its likelihood of causing impacts on online harms*."
- The NSPCC said: *"We agree with Ofcom's assessment of the causes and impacts of online child sexual abuse and exploitation (CSEA)."* And provided a range of additional evidence for our analysis; *"Below we highlight further evidence which should inform Ofcom and services' assessments of these harms. We would also like to direct Ofcom to the evidence review recently published by NSPCC, focusing on children's exposure to online sexual risks and the role technology plays in exacerbating or reducing these risks."*
- Innovate Finance said: "*Innovate Finance broadly agrees with Ofcom's assessment of the causes and impacts of online harms. We agree with the causes of these online harms which Ofcom has identified with regards to the proceeds of criminal offences, fraud and financial services offences*." The stakeholder also provided additional evidence for use in our analysis of the risk factors relating to fraud.

Our decision

A2.32 In the absence of any stakeholder feedback suggesting that we should take an alternative approach, we decided to maintain our overall method for collecting and analysing evidence and using this to create our Register of Risks as set out above in 'Methodology'.

Grouping of illegal harms

A2.33 Some stakeholders provided feedback on the way individual priority offences had been grouped in the kinds of illegal harm.

A2.34 In particular, the Global Alliance Against Traffic in Women (GAATW)[2387], [✂][2388] and the Scottish Government[2389] argued that unlawful immigration and human trafficking should have separate sections in Ofcom's Register of Risks. Each of these respondents highlighted that the harms are legally and factually distinct concepts which cannot necessarily be addressed using the same approach – largely because, among other things, who is targeted and affected by these offences is very different. Unlawful immigration is an offence against the state, while human trafficking is an offence against an individual.

A2.35 Stakeholders also raised question about the overlapping nature of the different kinds of illegal harm covered by the Act and subsequently grouped for the purpose of creating our Register of Risks. For example, Meta and WhatsApp[2390] highlighted that the explanation of CSEA offences, and in particular the grouping of Grooming and CSAM offences, would benefit from greater clarity: *"Ofcom's explanations in the Consultation of the different kinds of online harms could, at times, be clearer as to how the 'CSEA offences' (Child Sexual Exploitation and Abuse offences) category of harm is treated, and as to how risk ratings for this category and its 'CSAM' (Child Sexual Abuse Material) sub-category are to be derived."* Similarly, stakeholders raised concerns that some kinds of illegal harm appeared to 'overlap', and that one expression of risk on a service might constitute multiple kinds of illegal harm. This raised uncertainty about exactly how service risk assessment might work in practice.

Our decisions

A2.36 We have reviewed these kinds of illegal harm since the November 2023 consultation, drawing on consultation responses where relevant, and made two changes:

- We have separated the evidence on 'Human trafficking' and 'Unlawful immigration'. These are now addressed as separate kinds of illegal harm, each with a distinct chapter in the Register of Risks.
- We have separated the evidence on 'Encouraging or assisting suicide' and 'Encouraging or assisting serious self-harm'. These are now addressed as separate kinds of illegal harm, each with a distinct chapter in the Register of Risks. This decision was taken separate from stakeholder feedback and is due to encouraging and assisting serious self-harm not being classified as a priority offence in the Act, unlike the priority offence of encouraging or assisting suicide.[2391]

A2.37 The Risk Assessment Guidance for Service Providers contains additional clarifications regarding how CSEA harms should be assessed.

A2.38 Stakeholders had raised concerns about two or more harms being seen as 'overlapping'. This refers to where an offence or harm cannot be committed completely in isolation, and other offence(s) are legally inseparable from the first. The perceived implication of there being an overlap of this kind is that if service providers assess the risk of each kind of illegal harm independently, where they identify one, they logically must identify the other

[2387] GAATW response to November 2023 Illegal Harms Consultation, p. 1
[2388] [✂]
[2389] Scottish Government response to November 2023 Illegal Harms Consultation, p. 2
[2390] Meta and WhatsApp response to November 2023 Illegal Harms Consultation, p. 14
[2391] Assisting serious self-harm was expected to be a priority harm until late on in the drafting of our November 2023 Consultation, so the two were originally addressed together

overlapping risk. Where these harms are the only kinds present on a service that service would necessarily be classified as 'multi-risk'[2392], which has implications for the safety duties that apply. It is important to note this kind of overlap is distinct from where offences or kinds of illegal harm might often occur together, but remain legally distinct.[2393]

A2.39 We reviewed kinds of illegal harm where this overlap could potentially be occurring. Because the kinds of illegal harm have very different impacts and users affected may have very different needs, we are content that they should be risk assessed separately and considered separately for the purposes of the definition of multi-risk. To ensure we provide a comprehensive assessment of each kind of illegal harm covered by the Act within the Register of Risks, we have therefore retained separate chapters for all kinds of illegal harm.

A2.40 We have also added relevant chapters from the August 2024 Further Consultation, with 'Animal cruelty' and 'Obscene content showing torture of humans and animals (the s.127(1) offence)' now included in our Register of Risks.

A2.41 For our December 2024 Illegal Harms Statement, the priority illegal harms and relevant non-priority offences, including communications offences, for which we have chapters in the Register of Risks, are:

a) Terrorism
b) Child Sexual Exploitation and Abuse (CSEA):
    i) Grooming
    ii) Child Sexual Abuse Material (CSAM)
c) Hate
d) Harassment, stalking, threats and abuse
e) Controlling or coercive behaviour
f) Intimate image abuse
g) Extreme pornography
h) Sexual exploitation of adults
i) Human trafficking
j) Unlawful immigration
k) Fraud and financial offences
l) Proceeds of crime
m) Drugs and psychoactive substances
n) Firearms, knives and other weapons
o) Encouraging or assisting suicide
p) Foreign interference
q) Animal cruelty
r) Non-priority offence - Epilepsy trolling
s) Non-priority offence - Cyberflashing
t) Non-priority offence - Encouraging or assisting self-harm
u) Non-priority offence - False communications

[2392] In our November 2023 Consultation we proposed to define a service as multi-risk where it is high or medium risk for at least two kinds of illegal harms. This is important because some of the measures we are proposing target a wide range of online harms and we propose to apply the most onerous of these measures in our Codes only to services which are large and/or multi-risk.
[2393] For example, advertising a fake job could likely amount to both fraud and human trafficking offences, but we would still expect service providers to assess the risk of both kinds of illegal harm separately.

v) Non-priority offence - Obscene content showing torture of humans and animals (the s.127(1) offence)
w) Non-priority offence - Threatening communications
x) Search services
y) Governance, systems, and processes

New risk factors for certain illegal harms

A2.42 An important aspect of our evdience base is the extent to which it allows us to confidently determine and demonstrate links between harms and specific risk factors. We received welcome input and new evidence from a wide range of stakeholders demonstrating the links between risk factors and certain kinds of illegal harm that we had not previously been able to evidence as clearly as we are now. These are all addressed within the updated Register of Risks and detailed information can be found in the Annex 1: Further stakeholder responses. Examples of the kind of feedabck received include:

- Jonathan Hall KC, Independent Reviewer of Terrorism Legislation[2394] and the Children's Commissioner[2395] highlighted that 'age' as a risk factor for terrorism offences required further assessment and inclusion given the evidence that points to the potential vulnerability of children and young people to radicalisation online.
- The Victims' Commissioner for England and Wales[2396] highlighted that the draft Register of Risks did not suficiently explain the link between offline and online behaviour in cases of controlling and coercive behaviour.
- The Institute for Strategic Dialogue[2397] highlighted that video-sharing services have been identified as online spaces that can be used to commit or facilitate offences related to hate, targeting minorities and other protected groups, and provided new evidence not previously available for inclusion in the Register of Risks.

Our decision

A2.43 Where stakeholders provided new evidence on the risks of harm or raised points around a need for greater clarity or nuance in our assessment of the risks of harm, we have considered how best to take these into account. Where appropriate, we have updated the Register of Risks accordingly. This has led to hundreds of additions and every chapter has been updated in some way drawing on the feedback of stakeholders and wide variety of new evidence sources.

A2.44 In particular:

- We have added risk factors to most chapters, drawing in particular on new research and evidence that links particular functionalities and service types with kinds of illegal harm, ultimately providing a much more comprehensive evidence base than we had when publishing the draft Register of Risks in November 2023. For example, this includes adding network recommender systems as a risk factor in relation to harassment, stalking, threats and abuse; social media and messaging services as important service

[2394] Jonathan Hall, Independent Reviewer of Terrorism Legislation response to November 2023 Illegal Harms Consultation, pp. 1-6
[2395] Children's Commissioner response to November 2023 Illegal Harms Consultation, p. 20
[2396] Victims' Commissioner for England and Wales response to November 2023 Illegal Harms Consultation, pp.1, 4
[2397] Institute for Strategic Dialogue response to November 2023 Illegal Harms Consultation, pp. 3-5.

types in the commission of human trafficking offences; and gaming services in relation to proceeds of crime offences.

- We have added significantly to our commentary surrounding user base charactersitics across the Register of Risks, providing more information about who may experience a greater risk of harm. For example, including additional evidence in relation to a service's users who may be victims of intimate image abuse, foreign interference, or terrorism offences – where the age of a service's users has been added to ensure this is something that can be accounted for when assessing risks – among others. On the same theme, where appropriate, we have provided further information regarding user base characteristics that relate to perpetrators of offences online.

A2.45 We have updated our definitions and descriptions throughout, wherever a lack of clarity or specificity was causing confusion. For instance, we have updated our definitions of recommender systems to account for product recommender systems as well as content and network recommender systems.

A2.46 An explanation of the detailed changes we have made to the Register of Risks is set out at Annex 1: Further stakeholder responses. We have also noted where new risk factors were proposed, but which we have not added to the updated Register of Risks.

A2.47 The Risk Profiles have also been updated to accurately reflect this updated evidence base.

A2.48 We will continue to review new information as it becomes available and remain on the lookout for any characteristics that constitute new risk factors that should be reflected in our Register of Risks and/or Risk Profiles.

A3. Updating the Risk Profiles

Our proposals

A3.1 The Act requires Ofcom to prepare and publish 'Risk Profiles' based on the findings in our Register of Risks.[2398] The Risk Profiles are a vital tool to help providers understand the most important findings in our Register of Risks in a practical and meaningful way. Service providers must take account of our Risk Profiles when they conduct their risk assessments, and this will ensure they understand the characteristics of their services that may be particularly relevant to each kind of illegal harm. The Risk Profiles are crucial in translating the breadth and depth of the evidence in our Register of Risks for providers in a way that tangibly improves online safety for at-risk users.

A3.2 The Act gives Ofcom wide discretion about how we create the Risk Profiles. In particular, we can group services in whichever way we consider appropriate taking into account the **characteristics of services**, the **risk levels** and **other matters** identified in the Register of Risks.

November 2023 Illegal Harms Consultation

A3.3 In our November 2023 Consultation, we proposed an approach where Risk Profiles should be used to highlight:

- the **characteristics**[2399] **of services** that are likely to increase risk of harm (we refer to these as risk factors), and
- which **kinds of illegal harms** may be more likely to occur on their service as a result.

A3.4 We proposed presenting the Risk Profiles as tables with each row representing an individual risk factor (e.g. encrypted messaging or livestreaming). For each risk factor, we provided a high-level description of how the risk typically arises, and the illegal harms that are most relevant to that risk factor. **The tables do not set out all the risk factors from the Register of Risks.** They instead include those which **we have determined to be particularly important for services to consider.**[2400]

A3.5 We considered several options about whether we should separate Risk Profiles by U2U and Search, service-type, kind of priority illegal harm, or groupings of similar functionalities. **We opted to produce two Risk Profiles presented as two tables of risk factors; one for U2U services to consult ('U2U Risk Profile') and one for Search services to consult ('Search Risk**

[2398] Section 98(5) of the Act. The Register of Risks is Ofcom's own risk assessment of the impact of characteristics of services on the risks of harm to individuals from illegal content. For U2U services, this includes the risk of harm from the facilitation and commission of illegal harms, as well as users encountering illegal content. For Search, this includes only the risk of harm from users encountering illegal content. Details on our approach and our full findings from our risk assessment are available in the Register of Risks document.

[2399] Characteristics include a service's user base, business model, functionalities and any other matters we deem relevant to risk. Risk Profiles focus predominately on user base demographics, functionalities and business models. Step 2 of the risk assessment guidance provides information for services on user base size, governance, and systems and processes.

[2400] We recognised that Risk Profiles cannot fully capture the complexity and context of risk factors across all the harms considered.

Profile'). We proposed that services should consult the relevant table and decide which risk factors are relevant to them when doing their risk assessment.

- Some risk factors in the tables were those that only relevant service providers had to take account of in their risk assessment because they represent characteristics that only certain services have (for example, being able to 'comment'). We referred to these as **specific risk factors**, and service providers were expected to identify which of these apply to them. To help services do this accurately, we provided a list of Yes (Y) or No (N) questions, where each 'Y' answer corresponds to an additional risk factor in the tables.
- Some of the risk factors in the tables were things that all services must take account of such as **user base demographics**, **business model** and **commercial profile.** We referred to these as **general risk factors**. Given that there were only three general risk factors, we included high level information about all three in both the U2U and Search tables. We also provided information about different kinds of illegal harms where possible.

A3.6 When compiling the U2U Risk Profile, we reviewed the analysis set out in the Register of Risks and identified the specific risk factors which were most strongly associated with each illegal harm.[2401] We included these in the U2U Risk Profile. **We only included specific risk factors in the U2U Risk Profiles where the evidence connecting them to illegal harms was strong**. Where the linkages the Register of Risks identified between specific risk factors and illegal harms were less clearly evidenced, we excluded them from the U2U Risk Profile.

A3.7 However, we found there were fewer specific Search risk factors. This was because the range of characteristics on Search services was narrower than on U2U services, and there was less evidence available – including relatively limited information on the links between individual Search risk factors and specific kinds of illegal harms. **We therefore included all specific Search risk factors in the Search Risk Profile table and described the general risk of harm, rather than listing the key kinds of illegal harm for each risk factor.**

A3.8 We explained our view that this approach to Risk Profiles adapted well to each service and would result in a list of risk factors which were associated with an increased risk of harm for U2U and Search services.

A3.9 **We proposed that after consulting the relevant table, service providers should have identified the list of risk factors (and associated illegal harms) that apply to them, which they must take account of in their risk assessment**. We said this list will always include all general risk factors for either Search or U2U, plus any specific risk factors indicated by their answers.

A3.10 By taking account of our Risk Profiles in this way, we considered that services would have a good starting point for thinking about the level of risk their service may present for different kinds of illegal harm and which risk factors ordinarily contribute to that risk. More broadly, we explained service providers should use this information to help them assess

[2401] We determined that a qualitative methodology was better able to provide an accurate assessment of the evidence available given the complexity of the evidence and the lack of consistent or comparable numerical data across illegal harms. The methodology considered the strength of the evidence for different risk factors, common trends across illegal harms, and alignment with other aspects of our regulatory approach. For example, when considering "hyperlinks" as a risk factor, we considered how the evidence in the Register explained the relationship between hyperlinks and each kind of illegal harm individually, as well as considering the relationship between hyperlinks and illegal content more broadly. We also considered the relationship between hyperlinks and our wider regulatory approach, for example the Codes of Practice.

their risk level for each kind of priority illegal offence in Step 2 of the risk assessment process.

Alternative approaches to the Risk Profiles

A3.11 As set out in our November 2023 Consultation, we considered other ways to produce our Risk Profiles.[2402] These options included producing separate 'Risk Profiles' base on:

- different 'types' of service;
- each kind of priority illegal harm; and,
- functionality groups.

A3.12 We ultimately decided that our proposed approach of creating separate Risk Profiles for U2U and Search services was the most suitable option to meet our policy objectives.[2403] It enabled us to draw out the evidence in the Register of Risks robustly and accurately while also highlighting similarities in risk across different services and illegal harms. For example, we could explain how direct messaging, whether between buyers and sellers within an online marketplace, as a feature of a game, or on a private messaging service, could increase the risk of illegal harms. We expected this to help service providers interpret the evidence more broadly and think more systematically about the types of illegal harms that may occur on their service.

A3.13 Our proposed approach also presented our evidence in a way that was easy for service providers to use as they would only need to know what characteristics their service has to determine what risk factors we expect them to assess for different illegal harms. We also expected this structure to be able to be updated with limited burden on services, as we could easily add or remove risk factors, or edit the descriptions for existing risk factors as our understanding of the illegal harm changed.

August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty

A3.14 Our August 2024 Further Consultation set out several proposals that added animal cruelty and human torture content to our Codes of Practice and guidance. The Act made the animal welfare offence a priority illegal offence. We were unable to address it in our November 2023 Consultation proposals for the Risk Profiles as it was added at a relatively late stage in the legislative process.

A3.15 For the U2U Risk Profile, we proposed, in addition to our November 2023 Consultation proposals, to add animal cruelty as a key kind of illegal harm associated with risk factors 1a. Social media service, 1b. Messaging service, 4b. Group messaging, 5d. Commenting on content and 5e. Posting images or videos risk factors. We also proposed to add additional wording to the group messaging risk factor section to explain how the functionality may facilitate the illegal harm in manifesting. We did not propose any changes to the Search Risk Profile with regards to animal cruelty.

[2402] See Table 9.3 of the November 2023 Illegal Harms Consultation for further details on alternative options that were considered.

[2403] We considered each option against two main objectives: a) our approach should effectively present our evidence on what makes services risky; and b) our approach should be easy for all services to use.

Changes to U2U Risk Profile based on new evidence

A3.16 As described in our November 2023 Consultation, we conducted a qualitative analysis to identify which of the many specific U2U risk factors were most strongly associated with the certain kinds of illegal harm in our evidence base. We have repeated this process for the purposes of finalising our Risk Profiles in light of an expanded evidence base reflected in our Register of Risks following consultation.

A3.17 No additional risk factors have been added to either the U2U or Search Risk Profiles. However, we have added key kinds of illegal harm to several risk factors in the U2U Risk Profile. This has been mainly due to new evidence received from stakeholders on some kinds of illegal harm, but also owing to our August 2024 Further Consultation and changes to the grouping of priority illegal harms in the Register of Risks. For example, the key kinds of illegal harms associated with 1f. Marketplace and listing services now includes the drugs and psychoactive substances illegal harm whereas there were no changes to the key kinds of illegal harm associated with 3a. Fake user profiles. We have set out the changes in the table below.

Table 2. Changes to key kinds of illegal harms associated with specific risk factors in the U2U Risk Profiles

Specific risk factor	Key kinds of illegal harm in draft U2U Risk Profile	Changes to key kinds of illegal harm in final U2U Risk Profile
1a. Social media services	All illegal harms except firearms and other weapons offences	Now associated with all illegal harms.
1b. Messaging services	CSEA (grooming and CSAM), animal cruelty, controlling or coercive behaviour, drugs and psychoactive substances, unlawful immigration/human trafficking, proceeds of crime, fraud and financial services, and foreign interference offences.	Nearly all illegal harms except intimate image abuse and extreme pornography offences.
1c. Gaming services	Terrorism, CSEA (grooming), hate and harassment/stalking/threats/abuse offences.	No change.
1d. Adult services	CSEA (image-based CSAM), extreme pornography, and intimate image abuse offences.	No change.
1e. Discussion forums and chat rooms	CSEA (CSAM) and encouraging or assisting suicide or serious self-harm offences.	Terrorism, CSEA (grooming), foreign interference and intimate image abuse offences added. Encouraging or assisting serious self-harm offences removed[2404]

[2404] Please note, this is due to encouraging or assisting serious self-harm being a non-priority illegal harm. See paragraph A2.32 to A2.37 for further details.

Specific risk factor		Key kinds of illegal harm in draft U2U Risk Profile	Changes to key kinds of illegal harm in final U2U Risk Profile
1f. Marketplace and listing services		Terrorism, sexual exploitation of adults, firearms and other weapons and fraud and financial services offences.	Drugs and psychoactive substances and human trafficking[2405] added.
1g. File-storage and file-sharing services		Terrorism, CSEA (image-based CSAM) and intimate image abuse offences	No change.
2. Services which allow child users		CSEA (grooming and CSAM) offences.	No change.
3a. Services with user profiles	User profiles	CSEA (grooming), harassment/stalking/threats/abuse, drugs and psychoactive substances, unlawful immigration/human trafficking, and sexual exploitation of adults offences.	Fraud and financial services, proceeds of crime, foreign interference and hate offences added.
	Fake user profiles	CSEA (grooming), harassment/stalking/threats/abuse, controlling or coercive behaviour, proceeds of crime, fraud and financial services and foreign interference offences.	No change.
3b. Services where users can post or send content anonymously, including without an account		CSEA (CSAM), encouraging or assisting suicide or serious self-harm, hate and harassment/stalking/threats/ abuse offences.	Terrorism, foreign interference, drugs and psychoactive substances, firearms, knives and other weapons offences added. Encouraging or assisting serious self-harm offences removed.
4a. Services with user connections		Terrorism, CSEA (grooming), harassment/stalking/threats/abuse, controlling or coercive behaviour, drugs and psychoactive substances, fraud and financial services and foreign interference offences.	No change.

[2405] This is due to human trafficking and unlawful immigration now being treated as separate kinds of illegal harm. See A2.32 to A2.37 for further details.

Specific risk factor		Key kinds of illegal harm in draft U2U Risk Profile	Changes to key kinds of illegal harm in final U2U Risk Profile
4b. Services where users can form user groups or send group messages	User groups	CSEA (grooming), encouraging or assisting suicide or serious self-harm, drugs and psychoactive substances and unlawful immigration/human trafficking offences. (Given the similarity with group messaging, we would also suggested that service providers consider if any key illegal harms associated with that functionality are relevant to their service)	CSEA (CSAM), hate, fraud and financial services, controlling or coercive behaviour, foreign interference and intimate image abuse offences added. Encouraging or assisting serious self-harm offences removed.
	Group messaging	Terrorism, CSEA (CSAM), animal cruelty, intimate image abuse, and fraud and financial services offences. (Given the similarity with user groups, we would also suggested that service providers consider if any key illegal harms associated with that functionality are relevant to their service)	CSEA (grooming) offences added.
5a. Services with livestreaming		Terrorism, CSEA (grooming and image-based CSAM), encouraging or assisting suicide or serious self-harm, and harassment/stalking /threats/abuse offences.	Animal cruelty and hate offences added. Encouraging or assisting serious self-harm offences removed.
5b. Services with direct messaging		CSEA (grooming and CSAM), hate, harassment/stalking/threats/abuse, controlling or coercive behaviour, intimate image abuse and fraud and financial services offences.	Terrorism, unlawful immigration, human trafficking, proceeds of crime, and drugs and psychoactive substances offences added.
5c. Services with encrypted messaging		Terrorism, CSEA (grooming and CSAM), drugs and psychoactive substances, sexual exploitation of adults, foreign interference and fraud and financial services offences.	Unlawful immigration, human trafficking, and firearms, knives and other weapons offences.
5d. Services with commenting on content		Terrorism, animal cruelty, CSEA (grooming), encouraging or assisting suicide or serious self-harm, hate, and harassment/stalking/threats/abuse offences.	Fraud and financial services offences added. Encouraging or assisting serious self-harm removed.

Specific risk factor		Key kinds of illegal harm in draft U2U Risk Profile	Changes to key kinds of illegal harm in final U2U Risk Profile
5e. Services with posting images or videos		Terrorism, CSEA (image-based CSAM), animal cruelty, encouraging or assisting suicide or serious self-harm, controlling or coercive behaviour, drugs and psychoactive substances, extreme pornography and intimate image abuse offences.	Harassment/stalking/threats/abuse, human trafficking, unlawful immigration, hate and foreign interference offences added. Encouraging or assisting serious self-harm offences removed.
5f. Services where users can post or send location information		Harassment/stalking/threats/abuse and controlling or coercive behaviour offences.	CSEA (grooming) and human trafficking offences added.
5g. Services with re-posting or forwarding of content		Encouraging or assisting suicide or serious self-harm, harassment/stalking/threats/abuse, intimate image abuse and foreign interference offences.	Encouraging or assisting serious self-harm offences removed.
6. Services where users can post goods or services for sale		Drugs and psychoactive substances, firearms and other weapons, sexual exploitation of adults and fraud and financial services offences.	Human trafficking offences added.
7a. Services where users can search for user-generated content		Drugs and psychoactive substances, firearms and other weapons, extreme pornography, and fraud and financial services offences.	Terrorism and proceeds of crime offences added.
7b. Services with hyperlinks		Terrorism, CSEA (CSAM URLs) and foreign interference offences.	Fraud and financial services, encouraging or assisting suicide, and drugs and psychoactive substances offences added.
8. Services with recommender systems	Content recommender systems	Encouraging or assisting suicide or serious self-harm and hate offences.	Terrorism and foreign interference offences added. Encouraging or assisting serious self-harm offences removed.
	Network recommender systems	CSEA (grooming) offences.	Drugs and psychoactive substances offences added.

Stakeholder responses and decisions by theme

A3.18 We received several responses to our consultations regarding the Risk Profiles. There was support for our overall approach, but some responses argued for us to take a different approach on certain issues. The prominent themes that emerged included:

- Views on the proposed approach to and format of the Risk Profiles

- Whether the Risk Profiles should have more context specific information
- Suggested changes to the risk factors and key kinds of illegal harm in the Risk Profiles
- Responses regarding the Search Risk Profile
- Review of and updates to the Risk Profiles
- Other amendments to the Risk Profiles

A3.19 We have set out some of the key substantive points submitted by stakeholders under these themes and our responses below. Further responses have been set out in Annex 1: Further stakeholder responses.

Views on the proposed approach and format of the Risk Profiles

A3.20 Our proposed approach to the Risk Profiles in the November 2023 Consultation received broad support from a variety of stakeholders. For example, large services such as Microsoft stated that "*The Risk Profiles are clear and provide actionable guidance that will help regulated services understand and identify potential risks on their service*".[2407] Meta stated that the Risk Profiles were helpful and agreed that they were a useful starting point from which service providers can begin their risk assessments.[2408]

A3.21 Other large services such as LinkedIn and Match Group also responded favourably. LinkedIn stated that the Risk Profiles were a clear, thorough and practical resource.[2409] Match Group highlighted that our approach covered a broad range of risks and therefore was a good starting point.[2410]

A3.22 We also received positive feedback from other stakeholders such as Betting and Gaming Council, Evri, Mencap and Stop Scams UK mostly citing their support of our proposals.[2411]

A3.23 Similarly, our proposed amendments to the U2U Risk Profiles set out in our August 2024 Further Consultation also received broad support from many stakeholders. This included Blue Cross, Born Free Foundation, Dogs Trust, International Cat Care, RSPCA, Scottish SPCA, The Links Group and Wildlife and Countryside Link.[2412]

A3.24 Some stakeholders proposed that we should present the Risk Profiles in an alternative manner, such as by having separate Risk Profiles for different service types instead of only separate U2U and Search Risk Profiles. They suggested that this would make the Risk

[2407] Microsoft response to November 2023 Illegal Harms Consultation, p.6.
[2408] Meta response to November 2023 Illegal Harms Consultation, p.12.
[2409] LinkedIn response to November 2023 Illegal Harms Consultation, p.6.
[2410] Match Group response to November 2023 Illegal Harms Consultation, pp.5-6.
[2411] Betting and Gaming Council response to November 2023 Illegal Harms Consultation, p.4; Evri response to November 2023 Illegal Harms Consultation, p.3; Mencap response to November 2023 Illegal Harms Consultation, p.5; Stop Scams UK response November 2023 Illegal Harms Consultation, p.6.
[2412] Blue Cross response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.5; Born Free Foundation response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.2; Dogs Trust response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.2; International Cat Care response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.2; RSPCA response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.5; Scottish SPCA response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.5; The Links Group response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.4; Wildlife and Countryside Link response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.1.

Profiles clearer for service providers and allow nuance between risk factors associated with different service types by, for example, identifying how a particular risk factor can cause heightened risk of illegal harms occurring on one service type but not another.[2413]

A3.25 In contrast, the Centre for Competition Policy supported our decision not to present the Risk Profiles by service type. It stated that presenting it by service type may discourage service providers from considering how interlinkages between service types may increase or decrease levels of risk. It also stated that it would make it difficult for the Risk Profiles to adapt to new types of services that may combine or push the boundaries of the service type categories.[2414]

A1.1 The Scottish Government stated that it would be helpful if we provided Risk Profiles by 'sector type' to support non-commercial organisations that develop, licence, or run technology platforms. It explained that charity, community or voluntary organisations may need additional support as they are less likely to employ dedicated IT or risk professionals.[2415]

A3.26 Christian Action, Research and Education (CARE) was concerned that we discarded the option of designing the Risk Profiles by service type and that the considerations for our proposed approach included the level of burden for service providers. It stated that the overriding consideration for choosing an option should be the one that minimises harm regardless of burden on service providers.[2416]

A3.27 The British and Irish Law, Education and Technology Association stated that the Risk Profiles should be more 'granular' or 'unbundled' by service type, harms type or other relevant criteria and should include examples showcasing how, for example, a social media service would approach a risk factor associated with grooming.[2417]

A3.28 WeProtect Global Alliance queried whether our Risk Profiles would be formatted by sub-categories of illegal harms. For example, instead of just the general illegal harm of CSAM, we would have a granular breakdown including online grooming and AI generated CSAM with their associated risk factors. It stated that granular sub-categories would be preferable to accurately reflect the diversity of the harm types which each require different responses from industry players and manifest differently on services.[2418]

A3.29 Some stakeholders stated that the Risk Profiles were too broad, high-level and lacked nuance.[2419] Overall, they were concerned that services like theirs would be ascribed greater levels of risk, leading to disproportionate measures needing to be implemented, or that the Risk Profiles would not apply to them in the typical manner intended.

[2413] Airbnb response to November 2023 Illegal Harms Consultation, p.6; Booking.com response to November 2023 Illegal Harms Consultation, pp.3, 6.
[2414] Centre for Competition Policy response to November 2023 Illegal Harms Consultation, pp.11-13.
[2415] Scottish Government response to November 2023 Illegal Harms Consultation, p.4.
[2416] Christian Action, Research and Education (CARE) response to November 2023 Illegal Harms Consultation, pp.7-8.
[2417] The British and Irish Law, Education and Technology Association response to November 2023 Illegal Harms Consultation, pp.3-4.
[2418] WeProtect Global Alliance response to November 2023 Illegal Harms Consultation, p.7.
[2419] [✂]; Booking.com response to November 2023 Consultation, p.3; Reddit response to November 2023 Illegal Harms Consultation, p.20; Roblox response to November 2023 Illegal Harms Consultation, p.9; Safe Space One response to November 2023 Illegal Harms Consultation, p.6; techUK response to November 2023 Illegal Harms Consultation, p.14.

Our decision

A3.30 In our November 2023 Consultation, we considered various alternative approaches to producing Risk Profiles before proposing to separate them by U2U and Search services. The options we considered included producing Risk Profiles by service type, kinds of illegal harm or groupings of functionalities.[2420] We remain of the view that our proposed approach is comparatively better than the alternative ones in meeting our policy objectives.[2421] In summary, our proposed approach allows us to:

a) draw out the evidence in the Register of Risks robustly and accurately while also highlighting similarities in risk across different services and illegal harms.
b) present our evidence in a way that was easy for service providers to use as they would only need to know what characteristics their service has to determine what risk factors we expect them to assess for different illegal harms. Creating Risk Profiles that are easy to use makes it easier for service providers to do high-quality risk assessments.
c) update the Risk Profiles with limited burden on services, as we could easily add or remove risk factors, or edit the descriptions for existing risk factors as our understanding of the harm changes.

A3.31 In contrast, we found that producing Risk Profiles separated by service type did not allow us to effectively present our evidence on risk. This was because:

a) the evidence base for each service type was not robust or consistent enough to structure Risk Profiles around them alone. Some service types had considerable evidence, whilst others are limited, especially when assessing some kinds of illegal harm.
b) service definitions were not homogeneous or common – some are quite narrow, and several well recognised service types (for example, social media services and online adult services) contain a complex and wide range of characteristics, each with different levels of risk.
c) the evidence may define service types differently making it difficult to develop consistent definitions from our evidence base.

A3.32 We also considered producing a 'Risk Profile' for each kind of priority illegal harm. Services would have specific Risk Profiles each representing one kind of illegal harm to consult, for example an 'Intimate Image Abuse Risk Profile' or 'Terrorism Risk Profile'. However, we rejected this option because we considered it would not be easy for services to use. This was because:

a) services would need to identify which Risk Profile applied to them – to do this accurately, they would need some prior knowledge about the risks associated with each kind of illegal harm. This would be particularly hard for services with fewer resources or less internal expertise.
b) under our proposed approach, services only need to know their functionalities or features to identify the relevant information about risks. We expect this to be easier for all services to use.

[2420] See Table 9.3 of the November 2023 Illegal Harms Consultation for further details on alternative options that were considered.
[2421] We considered each option against two main objectives: a) our approach should effectively present our evidence on what makes services risky; and b) our approach should be easy for all services to use.

A3.33 Lastly, we considered producing separate Risk Profiles based on functionality groups. For example, we would have produced specific Risk Profiles covering 'user networking functionalities' (which would include user search and user connections) and 'user navigation functionalities' (which would include content search and hyperlinks).[2422] It would have allowed us to draw out links in the evidence across different types of illegal harms with regards to functionalities such as the sharing of CSAM and extreme pornography. However, we rejected this approach because:

a) it did not easily integrate the evidence on how individual functionalities within the same group may have different associations to either the same or different illegal harms. For example, livestreaming and commenting are both within the 'user communication' functionality group but have distinct links to risk throughout the Register of Risks.
b) it would not present a robust view on the evidence related to risk factors associated with user base, business model or other characteristics. This is because it would have to be structured through functionality groupings, and the evidence indicates that many of these risk factors interact with all functionality groups.

A3.34 Regarding CARE's response, we have a duty under the Act to consider the burden on services as part of our assessment of the proportionality of our overall approach. We consider the easier to use the Risk Profiles are, the more likely services are to complete their risk assessments properly and effectively. Therefore, creating user-friendly Risk Profiles is conducive to reducing harm. We have chosen the format which we consider most clearly demonstrates which risk factors give rise to a risk of specific illegal harms. By extension, we consider this format to be the most conducive to high quality risk assessments and effective harm reduction.

A3.35 We recognise that online risk is complex and nuanced. As WeProtect Global Alliance notes, harms manifest in numerous ways. In producing the Risk Profiles, we have had to strike a balance between bringing out the nuances associated with each of the kinds of illegal harms and creating a tool which is accessible and easy to use. We want to ensure that all services, including smaller and less sophisticated services can produce suitable and sufficient risk assessments. Therefore, we have created a simple tool which presents a relatively high-level summary of the linkages between risk factors and illegal harms. We acknowledge that this has resulted in us simplifying some of the nuance associated with the harms and creating broader Risk Profiles, as some stakeholders have pointed out, but this simplification does not prevent services from completing a suitable and sufficient risk assessment that meets their duties under the Act. As set out in our Risk Assessment Guidance, accounting for the Risk Profiles is only the first step in a four-step risk assessment process. Moreover, the Register of Risks provides more granular detail and nuance on illegal harms and risk factors. Where services need more detail than the high-level summary provided in the Risk Profiles, which we emphasise provide a non-exhaustive list of risk factors which we consider to be important, we would encourage them to look at the Register of Risks.

[2422] We retained aspects of this structure in our proposed approach. When presenting functionality-based risk factors, we organised them based on the groupings described in this option, which match those used in the Register (for example, grouping user identification factors together within the U2U Risk Profile).

A3.36 Overall, we consider that stakeholder responses have not provided us with sufficient evidence or reasoning to change our approach to the Risk Profiles. We have therefore decided to maintain the overall format and structure that we set out in our consultation.

Whether the Risk Profiles should have more context specific information

A3.37 Airbnb flagged that the risk factors identified by Ofcom would not always carry equivalent risk on different types of services. The result is that all services on which the proposed risk factors are present could automatically be seen as services on which certain harms are present, even when they are not.[2423]

A3.38 Reddit was concerned that its service offering meant that the Risk Profile framework would label it as higher risk of almost all illegal online harms, regardless of how these offerings are designed, deployed, or used. This, they argued, would result in a paperwork burden at the outset in a "guilty until proven innocent" approach, despite the unique aspects of its model that inherently reduces risk, such as its human-moderated, community-based structure and democratic approach to governance.[2424]

A3.39 Roblox felt that the Risk Profiles seem better suited and more relevant to traditional services as opposed to services like Roblox. It therefore considered that services should be free to determine a reasonable means of satisfying themselves that certain conclusions set out in the Risk Profiles do not apply to them without having to consult enhanced evidence inputs in every instance.[2425]

A3.40 Roblox also stated that the Risk Profiles should allow for context dependencies and recognise that there is at least some degree of flexibility for a provider to conclude that certain functionalities do not pose risks of harm. In other words, the existence of a functionality should not be determinative of its harm potential.[2426]

A3.41 Mobile Games Intelligence Forum felt that the presentation of gaming services as a risk factor for multiple illegal harms in the U2U Risk profile meant that all in-scope online games will be subject to the most burdensome multi-risk requirements unless they are able to show there is no evidence of each of those harms arising. They therefore wanted the risk assessment guidance to include consideration of the mitigation measures that these services have in place.[2427]

A3.42 techUK disagreed with the assumption that all file-sharing and file storage sites are "high-risk." It considered that categorising all file-sharing and file storage sites in the same manner is too simplistic and may result in a blunt instrument that fails to accurately represent the diverse nature of these services.[2428]

A3.43 Similarly, a service provider said that the Risk Profiles should consider the efficacy of a service's content moderation practices, business model and product functionality, and likelihood of harm occurring on the platform before concluding that file-sharing and file-

[2423] Airbnb response to November 2023 Consultation, p.5.
[2424] Reddit response to November 2023 Consultation, pp.24-25.
[2425] Roblox response to November 2023 Consultation, pp.9-11; [✂].
[2426] Roblox response to November 2023 Consultation, p.10; [✂].
[2427] Mobile Games Intelligence Forum response to November 2023 Illegal Harms Consultation, pp.1-3.
[2428] techUK response to November 2023 Consultation, p.14.

storage services pose an elevated risk of the illegal harms associated with each risk factor.[2429]

A3.44 Trust Alliance Group also stated that we should ensure services that have file-sharing functionalities, but are not a primary part of their service, are captured by question 1g of the U2U risk factor question list. This was so services which only provide this functionality as a supplementary feature would not be able to avoid related duties.[2430]

A3.45 UK Interactive Entertainment (UKie) stated that it would be disproportionate to equate the risk of CSAM appearing in video games with the risk of such content appearing on other online platforms, such as social media. This was because only 0.00025% of reports of CSAM or grooming material received by the National Centre of Missing and Exploited Children (NCMEC) were from video game platforms. It therefore felt that this reduced risk should be reflected in the Risk Profiles that apply to gaming services.[2431]

A3.46 Google proposed that 'service type' risk factors should be removed from the Risk Profiles. It listed a number of reasons including that service type definitions assume all services of a certain type have specific functionalities (leaving no nuance for adjustment in cases they do not) and that they are simply a combination of functionalities in which case it would be more logical to break a service down into its features and analyse those features regardless of the service types assigned.[2432]

A3.47 Google also stated that threshold for risk factors to be applicable to a service is too low, particularly because it seems that the existence of the functionality or feature on the service makes the risk factor applicable. Google proposed that the applicability of a risk factor should depend on a functionality or feature being used by a certain percentage of users or based on how significant an aspect of the service it was.[2433]

A3.48 Which? stated that we should consider the wider social aspects of a service in accordance with its risk of harm as part of the Risk Profiles. For example, it said that services like online dating services operate in a distinct social context, characterised by exchange of sensitive information and personal interactions, which make these services vulnerable to various forms of harm – particularly fraud and grooming – that are not currently accounted for.[2434]

A3.49 UK Safer Internet Centre proposed that we should consider the disproportionate risks faced by minorities and women online, as well as how extreme socioeconomic inequalities among platform users could facilitate grooming and sextortion, in the Risk Profiles.[2435]

A3.50 Cybersafe Scotland stated that it was important that children and young people from vulnerable backgrounds are separately identified and referenced for consideration within the Risk Profiles as they experience a significantly higher level of harm than is recorded in the draft Register of Risks.[2436]

[2429] [✂].
[2430] Trust Alliance Group response to November 2023 Illegal Harms Consultation, p.5.
[2431] UK Interactive Entertainment (Ukie) response to November 2023 Illegal Harms Consultation, p.3; UKie response to May 2024 Consultation on Protecting Children from Harms Online, p.12.
[2432] Google response to November 2023 Illegal Harms Consultation, pp.24-25.
[2433] Google response to November 2023 Consultation, pp.23-25.
[2434] Which? response to November 2023 Illegal Harms Consultation, p.3.
[2435] UK Safer Internet Centre response to November 2023 Illegal Harms Consultation, pp.4, 23.
[2436] Cybersafe Scotland response to November 2023 Illegal Harms Consultation, p.1.

Our decision

A3.51 A number of the responses suggested that our Risk Profiles were too mechanical, lacked nuance and did not take sufficient account of the idiosyncrasies of individual services or the steps they might already have taken to mitigate risks they face. We have had to strike a balance between capturing the nuances of the different ways in which harm occurs across a very diverse sector and producing a product that is easy to use. Overall, we consider that we have struck an appropriate balance.

A3.52 In this context, it is important to note that the Risk Profiles are one of a wide range of inputs service providers need to use when doing their risk assessment. As we explain in the Risk Assessment Guidance, service providers need to draw on a range of sources of evidence of risk on their service and use that evidence to assess all relevant aspects of their services, including existing steps they are taking to manage risks. The Risk Assessment Guidance makes clear that the Risk Profiles do not determine the level of risk for a particular service.

A3.53 For example, building on Google's point on thresholds, if only a very small number of a service's users used a particular functionality, and a service provider had accurate evidence to demonstrate this, then that may be a relevant consideration for a service provider to factor into its risk assessment. However, even if only a low proportion of users are interacting with a functionality, the service provider should still consider the likelihood and severity of harm this functionality is associated with. The same approach should be taken for Trust Alliance Group's point around file-sharing and file-storage functionalities even if it is not the primary functionality of the service. On balance, we therefore do not consider that we need to adjust the Risk Profiles in relation to this feedback.

A3.54 We have carefully considered UKie's concerns about the links the Risk Profiles make between CSEA and gaming services. We recognise that gaming services account for a materially smaller proportion of NCMEC reports than some other service types such as social media services. However, NCMEC reports are only one metric used to estimate the extent to which a service poses risks of CSEA. This is because specific CSEA offences, such as grooming, are significantly underreported and the quality of reporting may vary from service to service. At the same time, gaming services have a number of functionalities, which can increase the risk of CSEA occurring on a service. This can include, a large number of children using a gaming service and the ability to connect with people that they may not know. For these reasons, we consider there to be a risk of CSEA on gaming services and want to ensure that they are still captured within our Risk Profiles. We therefore consider that we do not need to adjust our Risk Profiles to reflect UKie's submission.

A3.55 In relation to Google's points, our identification of characteristics of a service that pose a risk of illegal harm as expressed in the Risk Profiles is based on extensive evidence on the causes and impacts of harm as presented in the Register of Risks. As such, the evidence points to some characteristics being inherently risky in relation to specific illegal harms and certain service types. Other relevant considerations that may impact overall risk of harm such as the number of users associated with a characteristic or the characteristic being a significant aspect of the service are accounted for in the wider service risk assessment process.

A3.56 The conclusion that those from minorities and women may face a disproportionate risk of harm online features prominently in our evidence base in the Register of Risks. While we

do not explicitly refer to minorities in the Risk Profiles, we do call out several protected characteristics such as race (including ethnicity) and religion under the general risk factors of user base demographics in the U2U Risk Profile, which would likely also cover ethnic or religious minorities. However, for the avoidance of doubt and to add further clarity to our Risk Profiles, we have decided to include direct mention of minorities.

A3.57 Similarly, we would in principle be open to including socioeconomic inequalities as an aspect of user base demographics which heightens risk. At this stage though, we are not aware of any evidence establishing the link between that and a higher risk of financially motivated sexual extortion or grooming.

A3.58 The user base demographics risk factor in the U2U Risk Profile already states that vulnerable users, particularly users with multiple protected characteristics, are more likely to experience harm from illegal content and are impacted differently by it. We consider Cybersafe Scotland's point is captured under user base demographics which, as a general risk factor, all service providers must take account of in their risk assessments. We therefore believe children and young people from vulnerable backgrounds would be considered in relevant risk assessments without the need for them to be separately identified in the Risk Profiles.

Suggestions for new risk factors for some illegal harm

A3.59 The FCA recommended that the fraud and financial services illegal harm should be added as a key kind of illegal harm to more risk factors in the U2U Risk Profile to ensure that it is appropriately highlighted when providers are conducting their risk assessments – which would facilitate awareness of the risks to their users.[2437] UK Finance also recommended adding the fraud and financial services illegal harm as a key kind of illegal harm to several risk factors.[2438]

A3.60 Samaritans suggested that it would be useful to add a risk factor around signposting or linking to online marketplaces and listings to purchase harmful items as this is an area of current concern from a suicide prevention perspective.[2439]

A3.61 Trust Alliance Group wanted us to include consideration of a service's 'business model' as a risk factor in the U2U Risk Profile based on its significance in shaping what a service looked like, how its users were protected and how those users understood their role within the service.[2440] It also asked for the inclusion of questions relating to 'audio messaging', 'AI generated content' and 'interoperability with other services' in question 5 of the U2U risk factor question list and hence for them to be risk factors in the U2U Risk Profile.[2441]

[2437] FCA response to November 2023 Illegal Harms Consultation, p.4. These risk factors included: 1e. Services with discussion forums and chat rooms, 3b. Services where users can post or send content anonymously, including without an account, 4b. Services where users can form user groups or send group messages, 5a. Services with livestreaming, 5g. Services with re-posting or forwarding of content, 7b. Services with hyperlinks, and 8. Services with recommender systems.
[2438] UK Finance response to November 2023 Illegal Harms Consultation, p.4. These risk factors included: 1e. Services with discussion forums and chat rooms, 3a. Services with user profiles, 3b. Services where users can post or send content anonymously, including without an account, 5a. Services with livestreaming, 5d. Services with commenting on content, and 7b. Services with hyperlinks.
[2439] Samaritans response to November 2023 Illegal Harms Consultation, p.4.
[2440] Trust Alliance Group response to November 2023 Consultation, p.2.
[2441] Trust Alliance Group response to November 2023 Consultation, p.7.

A3.62 Battersea Dogs & Cats Home, Born Free Foundation, Cats Protection and Social Media Animal Cruelty Coalition (SMACC) requested the addition of animal cruelty as a key kind of illegal harm to several risk factors in the U2U Risk Profile.[2442]

A3.63 International Cat Care, RSPCA and Scottish SPCA asked for animal cruelty to be added as a key kind of illegal harm to the adult services risk factor. This was because they have seen animal cruelty content being advertised on online adult sites including OnlyFans where acts of bestiality and torture of animals were being requested, though they acknowledged that this was rare on closed OnlyFans sites.[2443]

A3.64 In addition, the Scottish SPCA and International Cat Care also suggested that animal cruelty should be added as a key kind of illegal harm for the gaming services risk factor. They cited research that suggested that the types of games that children and young people play often have a bearing, albeit a modest one, on whether they are more accepting of cruelty towards animals.[2444]

A3.65 Battersea Dogs & Cats Home also asked for clarity as to why we chose not to include risk factors that have more limited evidence. It explained that their inclusion would ensure service providers are aware of the possibility that there may be animal cruelty content on their service and therefore make them more likely they to look for and recognise content that is illegal.[2445] Similarly, Cats Protection was unclear why we did not propose to include animal cruelty as a key kind of illegal harm for other risk factors (aside from those in our proposals) in the U2U Risk Profile. [2446]

A3.66 Born Free Foundation also requested the inclusion of animal cruelty offences in the Search Risk Profile given that animal cruelty content is likely to be findable via online search facilities. [2447]

A3.67 South West Grid for Learning (SWGfL) requested the addition of smaller services as a risk factor in the Risk Profiles. Its reasoning was that smaller services harbour concentrated communities engaging in illegal activities and that our current approach regarding the size of services excludes them in certain cases. It also suggested we add further details on how recommender systems might contribute to the spread of animal cruelty content.[2448]

[2442] Battersea Dogs & Cats Home response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, pp.5-8; Born Free Foundation response to August 2024 Further Consultation, p.2; Cats Protection response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, pp.6-8; Social Media Animal Cruelty Coalition (SMACC) response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, pp.2-3. These risk factors included varying combinations of: 1d. Adult services, 1e. Discussion forums and chat rooms, 1f. Marketplace and listing services, 1g. File-storage and file-sharing services, 3b. Services where users can post or send content anonymously, including without and account, 4a. Services with user connections, 5a. Services with livestreaming, 5b. Services with direct messaging, 5c. Services with encrypted messaging, 5g. Services with re-posting or forwarding content, 6. Services where users can post goods or services for sale, 7a. Services where users can search for user-generated content, 7b. Services with hyperlinks, and 8. Services with recommender systems.

[2443] International Cat Care response to August 2024 Further Consultation, p.4; RSPCA response to August 2024 Further Consultation, p.7; Scottish SPCA response to August 2024 Further Consultation, p.6.

[2444] International Cat Care response to August 2024 Further Consultation, p.4; Scottish SPCA response to August 2024 Further Consultation, p.6.

[2445] Battersea Dogs & Cats Home response to August 2024 Further Consultation, p.5.

[2446] Cats Protection response to August 2024 Further Consultation, p.5.

[2447] Born Free Foundation response to August 2024 Further Consultation, p.2.

[2448] South West Grid for Learning (SWGfL) response to August 2024 Illegal Harms Further Consultation on Torture and Animal Cruelty, p.3.

Our decision

A3.68 As we explain above, when determining what to put in the Risk Profiles we have had regard to the strength of the association between the risk factors under consideration and the harms. In general, we only include factors in the Risk Profiles where there is a strong association between risk factors and particular kinds of illegal harm. We have considered the stakeholder responses above carefully with this test in mind.

A3.69 Having regard to this test and the stakeholder submissions we received, we have added fraud and financial services as a key kind of harm to several risk factors in the U2U Risk Profile based on our qualitative analysis. These are:

- 3a. User profiles
- 4b. User groups
- 5d. Commenting on content
- 7b. Hyperlinking

A3.70 However, there were some suggested risk factors that we did not add fraud and financial services to as a key kind of illegal harm in the U2U Risk Profile at this point in time. These are:

- 1e. Discussion forums and chat rooms
- 3b. Posting or sending content anonymously
- 5a. Livestreaming
- 5g. Re-posting or forwarding of content
- 8. Recommender systems

A3.71 We thought that including the user groups functionality best captures a breadth of scenarios that could also occur on discussion forums and chatrooms. We also consider that for fraud, the ability to create fake user profiles is both an overlapping and a better evidenced vector of harm compared to the ability to post content anonymously.

A3.72 We have not seen sufficient evidence to warrant listing livestreaming, reposting or forwarding of content, or the existence of recommender systems as key risk factors for fraud and financial services offences in the Risk Profiles. In reaching this conclusion we have sought to strike a balance between ensuring that the Risk Profiles are sufficiently comprehensive and ensuring that they are selective enough to remain a useful analytical tool for identifying which services pose material risks.

A3.73 Having reviewed stakeholder responses and considered the available evidence, we have decided to add animal cruelty as a key kind of illegal harm to livestreaming in the U2U Risk Profiles. We consider that the evidence we have seen and stakeholder submissions are indicative that livestreaming functionalities are used to produce content which amounts to animal cruelty offences with sufficient frequency to warrant this change. As a result, the final risk factors for which animal cruelty is a key kind of illegal harm are:

- 1a. Social media services
- 1b. Messaging services
- 4b. Group messaging

- 5a. Livestreaming
- 5d. Commenting on content
- 5e. Posting images or videos

A3.74 However, the evidence we have analysed does not indicate that the other risk factors mentioned in the stakeholder responses referred to above are linked strongly enough with animal cruelty to justify amending our Risk Profiles.

A3.75 In relation to Samaritans' point about linking to online marketplaces, we do not consider that there is a need to add a new risk factor. In our view, the risk they have flagged is already reflected by the inclusion of the more general references to hyperlinking which appear in the Risk Profiles for which encouraging or assisting suicide offences are a key kind of illegal harm.

A3.76 We agree with the Trust Alliance Group's view that it is important that providers consider their services' business model when doing their risk assessments. This is already reflected in the Risk Profiles, and we have therefore not made any amendments in this regard. We have reviewed our evidence base and the available secondary literature and have not found sufficient evidence to justify including the other risk factors it referenced in the Register of Risk or the Risk Profiles. We will keep this decision under review as the evidence base evolves.

A3.77 Regarding Battersea Dogs and Cats Home's point about the inclusion of risk factors with only limited evidence, we explain that association of an illegal harm with a risk factor in the Risk Profiles requires strong evidence as a major consideration of our qualitative analysis of specific risk factors. In producing the Risk Profiles, we have had to strike a careful balance between being comprehensive and being selective. If we are insufficiently comprehensive, there is a risk that providers will fail to have regard to risk factors that have an important impact on the likelihood of harm occurring on their service. This would detract from the quality of their risk assessments. Set against this, if we included risk factors in the Risk Profiles in circumstances where there was only a weak link between the risk factor in question and the illegal harm, then the Risk Profiles would cease to provide a good basis for making judgments as to which harms a service should be most concerned about. This would reduce the quality of risk assessments conducted by service providers and could ultimately impair their ability to judge how best to protect their users.

A3.78 Regarding Born Free Foundations point, we do not list the key kinds of illegal harm for each risk factor in the Search Risk Profile. This is because there is less evidence available regarding the links between individual Search risk factors and specific kinds of illegal harms. However, we have acknowledged in Part 2 of the Register of Risks that it is likely that animal cruelty content will be available via search facilities. This is because there is some evidence for this type of content existing online, including in publicly available spaces on U2U services and stored on file-sharing and file-storage services.

A3.79 We have not made any amendments to our Risk Profiles to reflect SWGfL's submission. We consider that the Risk Profiles already take sufficient account of the way risks manifests on small services and the need to them to assess these risks robustly. We also consider we have set out how both content and network recommender systems can be used exacerbate the harm from illegal content more generally and in association with relevant

illegal harms. We have not referenced animal cruelty specifically as it has not been identified as a key kind of illegal harm for either of them.

Search Service Risk Profile

A3.80 DuckDuckGo stated that we failed to distinguish the nature of risk between general and downstream general search services in the Search Risk Profile which may lead to downstream general search services having a higher risk level and compliance burden [✂]. It also recommended that the specific and general search risk factors in the Search Risk Profile are adjusted to ensure that downstream general search services do not fall into the "specific risk" or "multi risk" categories in the Codes of Practice. [✂][2449]

A3.81 Skyscanner and Mid Size Platform Group were concerned by the lack and vagueness of the information provided on user base demographics for Search services in the Register of Risks and the Risk Profiles, particularly in relation to guidance on how search services are expected to consider the demographics of their user base. They sought clarification on how a service that does not collect such data, or has a majority of its users that do not create accounts, can do this.[2450]

A3.82 Skyscanner also recommended that the Illegal Harms and Childrens Search Risk Profiles make clear that the search prediction or suggestion functionalities risk factor is only related to free-form text predictive search, rather than more limited predictive search functionalities found on its vertical search service. Its concern was that though they have a predictive search function it is very limited and therefore the risk factor should not be applicable to it.[2451]

Our decision

A3.83 We acknowledge DuckDuckGo's comments and have therefore provided further information regarding our approach to downstream general search services and how they should undergo the risk assessment process in the Search Risk Profile. This includes additional commentary to clarify how the Search Risk Profile applies to downstream general search services[2452] and signposting to appropriate sections of this statement, such as Our approach to developing Codes measures, to ensure the necessary information can be located.

A3.84 In relation to comments regarding the vagueness of information, the purpose of the Risk Profiles is to guide services to understand what risk factors may be associated with a higher risk of illegal harm when they have relevant information. We provide further support in the Risk Assessment Guidance regarding the kinds of evidence inputs service providers may wish to consult to assess the level of risk of harm on their service, including a non-exhaustive list of evidence inputs relating to userbase. We consider that this guidance provides sufficient information to address Skyscanner's point.

[2449] DuckDuckGo response to November 2023 Illegal Harms Consultation, pp.2-3; [✂].
[2450] Mid Size Platform Group response to November 2023 Illegal Harms Consultation, p.4; Skyscanner response to November 2023 Illegal Harms Consultation, p.9; Skyscanner response to May 2024 Consultation on Protecting Children from Harms Online, pp.9-10.
[2451] Skyscanner response to November 2023 Consultation, p.10; Skyscanner response to May 2024 Consultation, p.6.
[2452] See the Search Risk Profile in the Risk Assessment Guidance and Risk Profiles regulatory document. For example, we have added a clarificatory footnote at Figure 2 Question 1 to explain how downstream general search services should use the Search Risk Profile.

A3.85 When looked at holistically, we consider that our Risk Assessment Guidance, Risk Profiles and Register of Risks already adequately reflect the points Skyscanner makes regarding search prediction. We have set out very clearly in these documents that we generally consider vertical search services to be low risk, and this is reflected in our Codes of Practice.

Review of and updates to the Risk Profiles

A3.86 Trustpilot queried the frequency with which Ofcom will update the Risk Profiles. It suggested that these updates should be grouped together and come into effect annually, which would be less burdensome on providers than doing multiple reviews per year. It also sought clarification around the timeframe in which providers would need to incorporate any updates to the Risk Profiles within their risk assessments. It suggested that it would be appropriate to mirror the three-month period given for complying with the risk assessment guidance.[2453]

A3.87 Similarly, in response to the May 2024 Consultation, techUK requested that we refrain from making frequent, significant changes to the Children's Risk Profiles to ensure that service providers are not required to continually update their compliance mechanisms, resulting in unnecessary costs and complexity. It suggested, for example, that Ofcom should not make significant changes to the Children's Risk Profiles within a year of publication.[2454]

A3.88 Snap stated that it would be helpful if we provided further details on the situations in which providers would need to update their risk assessments due to a significant change in the Risk Profiles as well as the timelines in which they would need to do this.[2455]

A3.89 Lloyds Banking Group suggested that Ofcom should review the Risk Profiles every six months with scope for event driven reviews to quickly incorporate emerging risks.[2456] UK Finance echoed these comments explaining that this would ensure that the update and review process remained "nimble" against evolving criminal methodologies.[2457]

A3.90 INVIVIA also said that the Risk Profiles and assessment methodologies should be regularly updated to reflect new types of risks and emerging online harms.[2458]

A3.91 5Rights Foundation stated that the Risk Profiles will struggle to keep up with emerging risks, such as generative AI and risk associated with its use to produce CSAM, if they are not updated regularly.[2459]

A3.92 Yoti said it would like to see more information as to how Ofcom will communicate with providers about the changes it may make to the Risk Profiles. It suggested that such communications should not be restricted solely to providers in-scope of the Act and be made available to the whole online safety ecosystem.[2460]

[2453] Trustpilot response to November 2023 Illegal Harms Consultation, pp.4-5.
[2454] techUK response to May 2024 Consultation on Protecting Children from Harms Online, p.9. We consider this response is relevant to our approach to reviewing and updating the Illegal Harms Risk Profiles.
[2455] Snap response to November 2023 Illegal Harms Consultation, p.7.
[2456] Lloyds Banking Group response to November 2023 Illegal Harms Consultation, pp.4-7.
[2457] UK Finance response to November 2023 Consultation, p.6.
[2458] INVIVIA response to November 2023 Illegal Harms Consultation, p.7.
[2459] 5Rights Foundation response to November 2023 Illegal Harms Consultation, p.14.
[2460] Yoti response to November 2023 Illegal Harms Consultation, p.7.

Our decision

A3.93 We recognise stakeholders' desire for clarity and certainty on the frequency with which we will be updating our Risk Profiles, how we will communicate these changes, and our expectations on when services would be expected to reflect that in their risk assessments. However, we have yet to decide the process and frequency of updates to the Risk Profiles. We will consider this in due course and have regard to the feedback we received from stakeholders.

A3.94 To clarify, service providers are not necessarily required to carry out a new full risk assessment when the Risk Profiles are updated. Instead, service providers need to take whatever steps are appropriate to keep their risk assessment up to date in light of the changes to the Risk Profiles as and when necessary, and if the update is relevant to them. This is set out in the Act.

A3.95 Furthermore, the frequency with which we update Risk Profiles will depend on changes to the underlying evidence base, including emerging risks we may need to reflect promptly, and as such it is difficult to determine this in advance. When we update the Risk Profiles, we will consider what that means for service providers needing to update their risk assessments and how best to communicate this. Again, this will need to be considered in all the circumstances at the relevant time.

Other amendments to the Risk Profiles

A3.96 Google stated that we had provided a different threshold in question 2 of the U2U Risk Profile question list than provided in the Act in the context of Children's Access Assessments. It asked us to clarify whether the policy intention was to create a separate threshold. If not, it suggested that ensuring the Risk Assessment Guidance aligned with the statutory definition would be helpful to avoid disparity with the Act.[2461]

A3.97 Trust Alliance Group also told us that question 2 of the U2U Risk Profile question list was too imprecise and could facilitate the under-reporting of children's presence on online services. It considered that it was well understood that children are present on services where they are not technically allowed to be and argued that the phrasing of the question could be improved. It proposed that either a supplementary request for evidence is included in the original question or the question is reframed to ask if service providers prevent child users from accessing some or all of their service, including a request for information relating to how this is achieved and what portions of the service are covered.[2462]

A3.98 Canadian Centre for Child Protection stated that children, if not effectively restricted from accessing adult services, could be harmed from exposure to the material.[2463] As a result, they said that that an additional risk could be highlighted in the adult services risk factor description.

A3.99 Dr Sandy Schumann proposed that 'search' functionality should be added to the U2U Risk Profile as a risk factor. They cited evidence that the ability to search, which is available on

[2461] Google response to November 2023 Consultation, pp.23. Figure 1, Question 2: Does your service allow child users to access some or all of your service?
[2462] Trust Alliance Group response to November 2023 Consultation, pp.6-7.
[2463] Canadian Centre for Child Protection response to November 2023 Illegal Harms Consultation, p.10.

many U2U services, for terrorist content enhances the risk of harm and the severity of potential impacts associated with terrorism content online.[2464]

A3.100 Trustpilot also recommended that we amend the title of section 8 of the U2U Risk Profile table so it is clear we are only referring to network and content recommender systems. It argued that, as drafted, the title was ambiguous and could result in services without either network or content recommender systems having to give attention to this area when they are not the intended targets, arguably diverting resources away from tackling relevant areas of risk.[2465]

Our decision

A3.101 Regarding Google's point on seeking clarification about whether question 2 of the U2U Risk Profile question list should align with the Children's Access Assessment, we acknowledge that additional clarity would help here as the two concepts are distinct. Question 2 is to help services include the risk of illegal harm to a child as part of their illegal content risk assessment. Children's Access Assessments are a new assessment that all regulated U2U and Search services must carry out to establish whether their service – or a part of it – is likely to be accessed by children. Services likely to be accessed by children will have additional duties to protect children online and they will also need to undertake a Children's Risk Assessment and implement safety measures to protect children online. As a result of this feedback, we have added a footnote to this question to clarify this point.

A3.102 In relation to this same question, we acknowledge Trust Alliance Groups comments and note that our own research produced insights showing that children under 13 years old readily used services, even where the terms of service did not allow them to do so.[2466] This shows that we would need to clearly set out that service providers must look at the reality of whether child users are accessing their service and not whether their terms of service allow child users. We have therefore amended the wording of question 2 in both Figure 1 and 2 of the Risk Assessment Guidance to reflect that we are referring to whether children are actually using the service – which is the determinant of risk.[2467] We have also amended the section 2 title of both the U2U and Search Risk Profile for clarity.[2468]

A3.103 We acknowledge Canadian Centre for Child Protection's point in relation to the risk to children from adult services. We have therefore made an addition the risk description box of 1d. Adult services in the U2U Risk Profile to highlight this point.[2469]

A3.104 In relation to Dr Sandy Schumann's point, we note that we already have a similar risk factor in the U2U Risk Profile called 'user-generated content searching' that we believe covers the 'search' functionality that they have suggested. We have added their evidence to the Register of Risks and now include terrorism as a key kind of illegal harm associated with 7a. User-generated content searching.

[2464] Dr Sandy Schumann response to November 2023 Illegal Harms Consultation, p.3.
[2465] Trustpilot response to November 2023 Consultation, p.8. Section 8 of the U2U Risk Profile table was titled 'Services with recommender systems'.
[2466] 'A window into young children's worlds', Ofcom, 2024. https://www.ofcom.org.uk/media-use-and-attitudes/media-habits-children/a-window-into-young-childrens-online-worlds. [accessed 1 September 2024].
[2467] Figure 1, Question 2: Do child users access some or all of my service? Figure 2, Question 2: Do child users access of my service?
[2468] U2U and Search Risk Profiles, section 2: Services which are accessed by child users.
[2469] "Furthermore, children could be harmed from exposure to this material if they are not effectively restricted from accessing these services."

A3.105 We have amended the title of section 8 of the U2U Risk Profile table to 'Services with content and/or network recommender systems' to align with the wording in our respective question and address Trustpilot's concerns.[2470]

Conclusion

A3.106 Having reviewed all consultation responses relating to our Register of Risks and Risk Profiles, we have decided to broadly proceed with our proposed approach for both the Register of Risks and Risk Profiles.

A3.107 Our Register of Risks has now been expanded and bolstered by hundreds of new pieces of evidence submitted to us in response to our consultations. We are grateful for the depth and breadth of relevant and high-quality research that has been brought to our attention. We have reflected the vast majority of these suggestions in our final Register of Risks.

A3.108 This has meant that we now have new evidence linking kinds of illegal harm to some risk factors that we did not have before. This has in turn led to the most notable change to the Risk Profiles where new kinds of illegal harm have been added to most of the U2U specific risk factors.

A3.109 As set out in the Risk Profiles section above, there were some legitimate concerns and requests regarding various elements of the Risk Profiles. Where appropriate, we clarified our approach or addressed these concerns by making changes to the Risk Profiles. For example, we clarified that the Risk Profiles are just one of several inputs that service providers need to consider when assessing risk. They provide a guide as to what characteristics can be risky based on the evidence gathered in the Register of Risks to help service providers conduct their risk assessments. However, as the Risk Assessment Guidance makes clear, services can and should consider a range of other factors including wider contextual factors alongside the Risk Profiles when doing their risk assessments.

A3.110 We have also added further information regarding our approach to downstream general search services and how they should undergo the risk assessment process in the Search Risk Profile. This includes additional commentary to clarify how the Search Risk Profile applies to downstream general search services and signposting to appropriate sections of this Statement, such as our chapter on 'Our approach to developing Codes measures', to ensure the necessary information can be located.

A3.111 Overall, we consider that our approach to the Risk Profiles works well to highlight key relevant findings from the Register of Risks. It is a crucial starting point for service providers to conduct their four-step risk assessments and will continue to be a valuable resource for service providers.

[2470] Question 8: Does my service use content or network recommender systems?