Protecting people from illegal harms online

Volume 2:
The causes and impacts of online harm

Consultation

Published: 9 November 2023
Closing date for responses: 23 February 2024

Contents

Section

Annex

4. Introduction to volume 2

This volume presents our assessment of the causes and impacts of illegal online harms based on the evidence that we have gathered over the past three years. The analysis we set out here forms part of our duty under the Act to assess the factors that can cause a risk of harm to individuals on a service. We expect services to have reference to it when they carry out their own risk assessments. Our assessment focuses on the over 130 priority offences defined in the Act. For ease of navigation, we have grouped these into 15 broad kinds of illegal harm. These include illegal harms such as: Child Sexual Exploitation and Abuse (CSEA), terrorism, fraud and hate speech, as well as the newly created Foreign Interference Offence, which addresses malicious online activity conducted by foreign powers (for example, state-sponsored disinformation campaigns). We summarise the findings of this assessment below and set out the detailed analysis in the body of this volume.

The illegal harms we have looked at are widespread and, in many cases, growing in prevalence. For example, 87% of adult internet users report having encountered a scam or fraud online and 25% of these people have lost money as a result. Almost a fifth of children experience sexual solicitation from adults they have chatted with online, and there was a 707% increase in the number of Uniform Resource Locator (URLs) which contain Child Sexual Abuse Material (CSAM) reported to the IWF between 2014 and 2021.

Online harms affect a large proportion of people in the UK and a wide cross-section of society. 63% of UK internet users say they have encountered potentially harmful content online in the past four weeks. However, children and people with certain protected characteristics are most likely to be affected. For example, 16% of minority ethnic internet users have encountered 'hateful, offensive or discriminatory content', compared to 11% of all internet users. Similarly, studies have shown that women are five times more likely to be victims of intimate image abuse. The more protected characteristics someone has, the more at risk of harm they are from priority illegal harms in the Act.

The impact of the harms we have looked at can be extremely severe. It is not limited to the online world but can also profoundly affect people's lives offline. A particularly clear example of this is online grooming, which can result in contact sexual abuse and cause lifelong negative psychological impacts including, loss of confidence, aggression, feelings of self-blame and lack of personal trust, as well as an increased risk of self-harm. In most cases the harms we have looked at primarily affect the individual experiencing them. However, in some cases they have a wider impact on society as a whole. For instance, state-sponsored disinformation campaigns can erode trust in the democratic process. All this underlines the need for the new legislation and shows that, while many services have made significant investments in tackling online harm in recent years, these have not yet been sufficient.

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, adult services, and file-storage and file-sharing services are all used to disseminate some of the types of harmful content we have looked at in this volume. Bad actors use both large and small services to spread illegal content, although the way in which they use large services sometimes differs from the way in which they use small services. 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.

Related to this, offenders often rely on multiple different types of service 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, encrypted 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 long tail of smaller services in scope of the Act.

Although a very wide range of service types pose risks of the priority illegal harms in the Act, certain service types appear to play a particularly prominent role in the spread of priority illegal content. In particular, our analysis suggests that file-storage and file-sharing services and adult services pose a particularly high risk of disseminating CSAM, and social media services play a role in the spread of an especially broad range of illegal harms. Similarly, certain ‘functionalities’ stand out as posing particular risks:

- **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**: There is 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 behaviour with reduced fear of the consequences. 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. For instance, livestreaming can be used as a way of conducting child sexual abuse by proxy, where children are coerced into abusing themselves or other children in real-time on camera.
- **Recommender systems**: 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.

We expect services to think about these risk factors when doing their risk assessments (see Volume 3). As we explain in Volume 4, we have designed a number of the measures in our Codes of Practice to target high-risk service types and functionalities.

The functionalities we describe above are not inherently bad and have important benefits. 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 role of the new online safety regulations is not to restrict or prohibit the use of such functionalities, but rather 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 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. Image-generation models, for example, can be used in some cases to create 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 how to make explosive materials.

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 arrangement 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 we take on our new role will therefore be ensuring that services do robust risk assessments and have appropriate governance arrangements in place. We discuss this further in Volume 3.

4.1 In this volume, we set out our understanding of the causes and impacts of online harm. We explain how we have compiled our evidence base for our sector-wide risk assessment, set out our key findings and detail the analysis of our sector-wide risk assessment in our 'Register of Risks'.

4.2 This volume is structured as follows:

- **Chapter 5, 'Evidence and methodology of our risk assessment'**: sets out the methodology for conducting our sector-wide risk assessment for illegal content. This sector-wide risk assessment 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 characteristics relevant to such risks of harm.
- **Chapter 6, 'Register of Risks':** sets out the draft analysis and findings of our sector-wide risk assessment to improve Ofcom's and services' understanding of risk.

4.3 We have produced detailed research and analysis in our draft Register of Risks (Register) to help services comply with their obligations under the Act. Our Register sets out our full risk assessment for illegal content on U2U and search services, and considers the use of regulated U2U services to commit or facilitate priority offences.

4.4 The draft Register is split into three parts:

a) The first part identifies characteristics of a U2U service that may lead to increased risks of harm to individuals: this includes its 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[1]. The risks from each kind of illegal harm on U2U services are explored in their own chapters.

[1] 'Kind of illegal harm' includes terrorism offences, Child Sexual Exploitation and Abuse (CSEA) offences, Encouraging or assisting suicide and serious self-harm, Hate offences, Harassment, stalking, threats and abuse, Controlling or coercive Behaviour, Drugs and psychoactive substances, Firearms and other weapons, Unlawful immigration and human trafficking,

b) The second part looks at the same characteristics of a service 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 the kinds of illegal harm together.

c) The third and final part looks at how a service's governance, systems and processes 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.

d) We also have an Annex which includes a glossary of terms used throughout the Register, as well as some further context to harm which we identified.

4.5 We are consulting on this draft Register of Risks and invite feedback on its clarity and ability to assist services to inform how they evaluate risks on their services. We have set out specific consultation questions in Chapter 6 in this volume on issues where we would particularly welcome feedback and further supporting information to inform our final version of this Register. See Consultation Annexes 1-4, for more information about how to respond to our consultation.

4.6 Having reviewed responses to this consultation, we will then publish our final decisions in a Statement and our final version of this Register of Risks.

Sexual exploitation of adults, Extreme pornography, Intimate image abuse, Proceeds of crime, Fraud & financial services, Foreign Interference Offence, Communications offences (false communication, threatening communication, epilepsy, cyberflashing).

5. Evidence and methodology for conducting our risk assessment

What is this chapter about?

This chapter explains how we have conducted the analysis of our sector-wide risk assessment (presented in the 'Register of Risks'), including the evidence used, the offences considered, and the risk factors analysed. This chapter also looks at the considerations involved in assessing the risk of harm to individuals.

Introduction

5.1 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.[2]

5.2 For this consultation, Ofcom has conducted a risk assessment for illegal content on U2U and search services, and considered the use of U2U services to commit or facilitate priority offences.[3] In Spring 2024, Ofcom will carry out a further risk assessment for content harmful to children.

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

5.4 Ofcom's Risk Profiles consider the characteristics of a service (including functionalities, user base and business model) that our risk assessment suggests may be relevant to the risk of certain kinds of illegal harm.

5.5 These Risk Profiles relating to risks of harm from illegal content are published as part the Service Risk Assessment Guidance (Consultation Annex 5), as services must take account of them when doing their own risk assessments.

5.6 Ofcom must keep both the Register and Risk Profiles up to date. We will monitor harms and regulated services trends and will revise our Register 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

[2] '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.

[3] See Volume 1, Chapter 2: Offences and Harms for more information about the priority offences.

may include technologies such as immersive online virtual worlds, augmented realities, and generative artificial intelligence ('generative AI').[4]

5.7 Our risk assessment of the sector aims to help services understand what increases the risks of illegal harm online. The variety of functionalities and offences considered has led us to identify an array of specific risks; however, it is important to note that service characteristics in themselves are not inherently harmful.

5.8 This chapter explains how we have conducted the analysis of our sector-wide risk assessment, including the evidence used, the offences considered, and the risk factors analysed. 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 to individuals.

5.9 The detailed findings of our risk assessment can be found in our provisional Register in Chapter 6 of this volume. The Risk Profiles can be found in Consultation Annex 5: Service Risk Assessment Guidance.

Methodology for the analysis

Evidence

5.10 Our risk assessment process has consisted in identifying and analysing a repository of quality assured evidence of nearly 1000 individual sources. We have considered responses from our July 2022 call for evidence, as well as relevant Ofcom research, academic papers from a range of disciplines, government bodies, 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 in this consultation, we have taken steps to ensure that our evidence sources are robust and reliable. In particular, we have considered the evidence by reference to the following criteria: method, robustness, ethics, independence and narrative.[5] [6]

5.11 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.

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

[4] We are aware of the debate around the potential risks that generative AI may pose. Given the pace of developments regarding generative AI and the fact that the evidence base in this area is still developing, we have considered this technology in a limited manner in this version of the Register. Our Register considers some of these risks.

[5] '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.

[6] 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.

5.13 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 helped understand online experiences, either by complementing any UK evidence available, or providing additional insights in cases where there was no UK evidence.

Evidence base

5.14 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.[7]

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

5.16 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 (e.g. extreme pornographic content offence), and for how services were used for the commission or facilitation of certain priority offences (e.g. unlawful immigration and human trafficking, and the sale of firearms/weapons). 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.

5.17 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. Due to the fast pace of technological changes and the speed at which risks of harm can manifest online, some of the evidence used within the 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.

5.18 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 to have more general application. Its inclusion should not be seen as a judgement about the online safety practices of those specific services. 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.

[7] 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.

Offences considered in Ofcom's risk assessment

5.19 The relevant offences considered in our risk assessment are:

a) **Priority offences** or **priority illegal content, which include** terrorism offences, offences related to CSEA and other priority offences. These are detailed in Consultation Annex 5: Service Risk Assessment Guidance. So-called '**inchoate offences**'[8] are also treated as priority offences.

b) **Relevant non-priority offences**[9] 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 or giving photograph or film of genitals ('cyberflashing'); and **self-harm offences**.

5.20 To make our assessment as accessible as possible, we have grouped these offences in the Register 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 the particular observation or evidence is relevant only to specific offences.

The kinds of illegal harm are:

- 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[10]
- 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

[8] As explained in Volume 1, Chapter 2: Offences and 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.

[9] 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.

[10] The new self-harm offence is not yet in force and is not a priority offence. However, we have included suicide and self-harm in the same kind of illegal harm. Most of our evidence base relates to both suicide and self-harm, so we have considered them together in our risk assessment. The evidence we analysed does not often distinguish between content that focuses solely on suicide, compared to content which focuses on self-harm that is potentially life-threatening. We also think that services may find it easier to consider these offences together, so have provisionally included them within the same kind of illegal harm.

- Intimate image abuse offences
- Proceeds of crime offences
- Fraud and Financial services offences
- Foreign Interference Offence
- Communication offence - *False communications offence*
- Communication offence - *Threatening communications offence*,
- Communication offence - *Offences of sending or showing flashing images electronically ('epilepsy trolling')*,
- Communication offence - *Offence of sending etc photograph or film of genitals ('Cyberflashing')*

5.21 At a fairly late stage in its consideration of the Bill which became the Online Safety Act, the offence in section 4(1) of the Animal Welfare Act 2006 (unnecessary suffering of an animal) was added to the list of priority offences. We will consult in due course on how we propose to include that offence in our Register.

Understanding service characteristics as risk factors

5.22 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 and Risk Profiles.

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

5.24 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 Table 1 at the end of this chapter. Where possible, we have also used these terms consistently across the other regulatory products included in this consultation.

5.25 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.

5.26 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 risky 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

5.27 Due to the nature of risk, we also distinguish two ways in which goods or services may be promoted on a service. This distinction was made because in some cases services are paid to promote content or 'advertisements' which represents a source of revenue. On the other hand, while users can promote goods and services by posting them for sale, in many cases the service is not paid to advertise them. The risks associated with how a service generates revenue differs from what functionalities are offered to users and how they might be used.

a) **Advertising** refers to paid-for advertising that generates direct advertising revenue for the service. This includes display advertising[11], classified advertising[12], and search advertising.[13] We cover this under 'business model'.[14]

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, and therefore it is not designed to generate direct advertising revenue for the service as classified advertising does. We cover this under 'functionality'.[15]

5.28 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.[16]

a) **Services with a large user base**: refers to services with a large number of monthly UK users

[11] '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).

[12] '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.

[13] '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.

[14] 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.

[15] This also sometimes considered under the umbrella of 'organic advertising'.

[16] 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.

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

Appendix

Table 1: Definitions for each characteristic

Characteristic	Description
Service type	e) 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.[17]
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[18] 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.

[17] 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).

[18] 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).

Characteristic	Description
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:[19] 1. Revenue models: how the service generates income or revenue e.g. through advertising, subscription, donation, transaction fees etc. 2. Growth strategy: how the service plans to expand its business. For instance, through growing revenue and number of users.
Commercial profile	f) 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[20] and rate of growth in relation to users or revenue.
Governance, systems and processes (GSP)	g) 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: **I)** **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. **II)** **III)** **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.

[19] '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.

[20] '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.

6. Introduction to Ofcom's Register of Risks for illegal content

6.1 Ofcom research has found that internet users think the benefits of being online generally outweigh the risks.[21] Despite this, we found that a majority of UK internet users have seen or experienced something potentially harmful online.[22] Following such exposure, users can suffer from a wide range of harms.

6.2 The purpose of the analysis in this Register of Risks is to understand the physical and psychological harm to individuals that can arise due to the risks of harm identified.[23]

6.3 This chapter gives an overview of the characteristics we assess in our risk assessment as relevant to the risks of harm to individuals.

Ofcom's approach to our risk assessment

Service characteristics

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

6.5 The **characteristics** of a service include any aspect of a service, including its functionalities, user base, business model, governance, and other systems and processes.[24]

- **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.[25]

[21] 70% of internet users 13 years old and over think benefits outweigh the risks. Source: Ofcom, 2022. Online Experiences Tracker. [accessed 10 September 2023].

[22] 63% of internet users 13 years old and over had seen or experienced something potentially harmful in the past four weeks.[Note: these may capture a broad range of potentially harmful experiences that go beyond illegal harms] Source: Ofcom, 2022. Online Experiences Tracker. [accessed 10 September 2023].

[23] 'Risks of harm' refers to the duties set out in the Act. For U2U services, risks of harm include those arising from (a) content on U2U services which amounts to the offences listed in Schedules 5, 6 and 7 of the Act; and (b) the use of these services for the commission and/or facilitation of these offences (collectively, the 'risks of harm'). For search services, 'risk of harm' refers to (a) content which amounts to the offences listed in Schedules 5, 6 and 7 of the Act.

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

[25] A non-exhaustive list of functionalities is provided in section 233 of the Act. 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

- **User base** refers to the users of a service. A user does not need to be registered with a service to be considered a user of that service.
- **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, we adopt a narrow definition that includes revenue model and growth strategy.[26] '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.
- **Governance, systems and processes** (GSP):

i) **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.

ii) **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.

6.6 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,[27] 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.

6.7 Within the analysis for the Register, where we found 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

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).

[26] '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.

[27] 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.

associated with a risk of one or more kinds of illegal harm.[28] For instance, direct messaging is a functionality that has been identified as a risk factor for some offences.

6.8 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.

6.9 Further information on this, including the full list of the most prominent, and potentially harmful, risk factors, is included in the Risk Profiles (Consultation Annex 5: Service Risk Assessment Guidance). More information and definitions of terms used throughout this Register can be found in the Glossary (in the Annex of this document, chapter 6V).

How risk factors associated to characteristics have been identified

6.10 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?[29]

6.11 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.

6.12 For instance, end-to-end encryption guarantees a user's privacy and security of messages, but makes it harder for services 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

[28] 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.

[29] 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.

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.

6.13 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.

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

Note on evidence

6.15 For information on how we collected the evidence and our evidence base, please refer to Chapter 5: Evidence and methodology of our risk assessment.

Part 1: User-to-User Services

6A. Introduction to U2U

Aim and structure

6A.1 This part of the Register presents a detailed analysis of the kinds of illegal harm, and their associated risks, on user-to-user (U2U) services. The risk factor analysis is summarised in Risk Profiles; services must take account of the Risk Profiles (Consultation Annex 5: Service Risk Assessment Guidance) when carrying out their own risk assessments. This Register, and the chapters within this part of it, can be referred to in order to assist with the assessment.

6A.2 This part includes 18 chapters which reflect the kinds of illegal harms[30] presented in Chapter 5: 'Evidence and methodology for conducting our risk assessment' of this volume. The chapters distinguish each priority offence where the evidence permits, and considers risks of harm to users more widely, where justified. For example, chapter 6B: Terrorism includes a number of priority offences relating to terrorism. Within it, we have evidence pointing to the risks of harm relating to a number of the terrorism offences listed, while other evidence may point to the particular harm of one terrorism offence in particular. We have also analysed evidence for other relevant non-priority offences (as detailed in Evidence and methodology for conducting our risk assessment, Chapter 5 of this volume) using the same structure as the other chapters.

6A.3 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.

6A.4 Each chapter is structured as follows:

a) Summary of the chapter, including the risk factors identified and those included in the Risk Profiles.
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

[30] 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 offences (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')). * The new self-harm offence is not yet in force and is not a priority offence. However, we have included suicide and self-harm in the same kind of illegal harm. Most of our evidence base relates to both suicide and self-harm, so we have considered them together in our risk assessment. The evidence we analysed does not often distinguish between content that focuses solely on suicide, compared to content which focuses on self-harm that is potentially life-threatening. We also think that services may find it easier to consider these offences together, so have provisionally included it here.

develop a better understanding of how specific characteristics relate to, and impact, the risks of harm.

U2U services

6A.5 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.

6A.6 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. 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 Sectors chapter (Volume 1, Chapter 3).

Service types

6A.7 Below are the service types we have considered in the following chapters. This is not an exhaustive list or a classification that sets expectations about a service's risk assessment. A U2U service may simultaneously include more than one service type, and some might also be a feature of a wider service.

a) **Social media services**: Social media services connect users and enable them to build communities around common interests or connections.

b) **Video-sharing services**: Video-sharing services allow users to upload and share videos with the public.

c) **Adult services**: Adult service services are primarily used for the dissemination of user-generated adult content.

d) **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.

e) **Marketplaces and listings services**: Marketplaces and listings services allow users to buy and sell their goods or services.

f) **Dating services**: Dating services enable users to find and communicate with romantic or sexual partners.

g) **Gaming services**: Gaming services allow users to interact within partially or fully simulated virtual environments.

h) **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.

i) **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.

j) **Information-sharing services**: Information sharing services are primarily focused on providing user-generated informational resources to other users.

k) **Fundraising services**: Fundraising services typically enable users to create fundraising campaigns and collect donations from users.

l) **Payment services**: Financial payment providers often have websites or applications that enable users to send and receive money.[31]

6A.8 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.

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

6B. Terrorism offences

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 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 on communities can be substantial, both physically and mentally.

Service type risk factors:

Terrorist content encountered by UK users do not rely on a single service but on many services and their associated functionalities. We have found that a lot of terrorism content is first posted on smaller U2U services and then linked to from larger, higher-reach services.

A wide range of types of U2U services are known to be used by terrorist actors. **Social media services** are particularly relevant to perpetration of this harm because of their reach and popularity. Terrorist content is also often identified on **file-storage and file-sharing services. 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. These types of service have therefore been included in the risk profiles.

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.

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, and the risks of harm to individuals exposed to this content, as it can lead to encouragement, incitement, the dissemination of material such as weapons training, and the recruitment of people. **User connections** also allow terrorism content to be disseminated through users' networks, especially when official pages or channels are removed. Due to their role in sharing illegal content, posting content and commenting on content, hyperlinking and user connections have been included in the risk profiles.

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. As a key risk factor for terrorism offences, livestreaming has been included in the risk profiles.

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. These two functionalities are included in the risk profiles due to their significance in propagating terrorism content.

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. Our evidence finds that **direct messaging** and **ephemeral messaging** is 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. Users have also been shown to create **fake user profiles** by altering their usernames to avoid having their account blocked.

Business model risk factors:

The **capacity and maturity** of a service can contribute to risk and be exploited by perpetrators, with varying impact on users. Low capacity and early-stage services may deal more with risks associated with their limited knowledge, resources or technical capability to moderate terrorism content. This reflects the opportunistic nature of perpetrators to use online services to heighten the risks of harm from terrorism, for varying purposes.

Introduction

6B.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the terrorism 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').

6B.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.

Relevant offences

6B.3 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to terrorism offences, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 5 of the Act.

6B.4 The priority offences for terrorism are the following:

- membership of a proscribed organisation[32]
- 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 uniform of 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
- publishing information about members of the armed forces etc
- inciting terrorism outside the United Kingdom
- use of noxious substances or things[33]
- encouragement of terrorism[34]
- dissemination of terrorist publications
- preparation of terrorist acts
- training for terrorism

[32] 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.

[33] The offence listed at point (q) refers to the offence under section 113 of the Anti-terrorism, Crime and Security Act 2001.

[34] 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.

- terrorist threats relating to radioactive devices etc.

6B.5 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6B.6 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** (see Volume 5, Chapter 26).

6B.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.[35]

6B.8 The UK Government publishes a list of proscribed terrorist groups or organisations[36] which include those proscribed by the UK Home Secretary[37] following assessment against a number of factors, including the specific threat they pose to the UK. 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 to professing to belong to, a proscribed organisation in the UK or overseas.

6B.9 Examples of terrorism content online could include 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, inviting support, and the promotion of terrorism content through discussion forums and chat rooms, as well as on social media services.

How terrorism offences manifest online

6B.10 This section is an overview which looks at how terrorism manifests online**,** and how individuals may be at risk of harm.

6B.11 To put the risks of harm to individuals from terrorism content into context, Ofcom's Online Experiences Tracker (OET) found that 3% 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 three-quarters (75%) expressing high concern.[38]

[35] The Crown Prosecution Service, 2022. Terrorism. [accessed 29 June 2023].

[36] The Home Office, 2021. Proscribed terrorist groups or organisations. [accessed 29 June 2023].

1.1 [37] The UK Security Service (MI5) is the primary agency responsible for counter terrorism operations in the UK. MI5 investigates threats from three ideological streams - International (such as Islamist groups Al-Qaeda or Islamic State of Iraq and Syria), right-wing, and left-wing, anarchist and single-issue terrorism (LASIT). Source: The UK Security Service, n.d. Counterterrorism [accessed 29 June 2023].

[38] . Comprises wave 1 and 2 combined data set. Question 7 '*Below is a list of things that someone may come across on the internet. Please tell me on a scale of 1 to 5, where 1 means 'mildly concerned' and 5 means 'very concerned', how concerned you are about the below existing online*.' was only asked of participants in wave 1. Source: Ofcom, 2023. Online Experiences Tracker 2021-2022. [accessed 4 October 2023].

Risks of harm to individuals presented by online terrorism offences

6B.12 Although the risk of coming across terrorism content is relatively small, there have been instances of this on U2U services, leading to a risk of harm to individuals. Our evidence shows that terrorist groups and actors use online services in specific ways, explored in greater detail below.

6B.13 Social media services are a tool for dispersing terrorism content and is a medium that terrorist organisations use to recruit, but there is analysis indicating that these are just one online arena that terrorist actors use.[39] Evidence points to terrorism content moving from larger, mainstream social media services to smaller services. A Tech Against Terrorism[40] 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"*.[41]

6B.14 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.[42] Tech Against Terrorism reported on this resurgence of terrorist-operated websites during 2021, explaining that terrorist organisations create these to further their goals.[43]

6B.15 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".[44] This tactic used by potential perpetrators makes it harder for a service to track down terrorism content, which can lead to it reaching a higher number of individuals.

6B.16 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'[45] shows that the internet[46] plays an increasingly

[39] 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].

[40] 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.

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

[42] 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.

[43] Tech Against Terrorism report, 2021.

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

[45] 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].

[46] 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

prominent role in radicalisation processes, in line with wider society's increased use of the internet.[47]

6B.17 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.

6B.18 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*".[48]

6B.19 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. Our evidence shows that explicit threats of violence, including in relation to the use of weapons, which can be disseminated online, can induce anxiety, fear or panic in a population or subset of the population.[49]

6B.20 For further evidence on the risks of harms to individuals that can occur from threats of violence, see chapter 6E: Harassment, stalking, threats and abuse.

Evidence of risk factors on user-to-user services

6B.21 We consider that the risk factors below are liable to increase the risks of harm relating to terrorism offences. This is also summarised in the grey box at the start of this chapter.

Risk factors: Service types

6B.22 Research indicates that a very broad range of types of U2U service 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.[50]

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].

[47] 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].

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

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

[50] Further information on the specific ways in which these services may be used can also be found under Risk factors: Functionalities and recommender systems.

Social media services and messaging services

6B.23 Our evidence highlights that many terrorist organisations or terrorist actors will use larger, mainstream social media services due to their substantial reach.[51] Tech Against Terrorism has reported that "terrorist networks are increasingly attempting to operate on mainstream social media platforms by masquerading as legitimate news organisations".[52]

6B.24 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.[53] 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.[54]

Discussion forums and chat room services

6B.25 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.[55]

File-storage and file-sharing services

6B.26 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".56

6B.27 Recent evidence suggests that the majority of terrorism content is identified on file-storage and file-sharing services. This is where most terrorism content was identified by Tech Against Terrorism. The use of file-sharing services is combined by 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.[57]

Gaming services

6B.28 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. [58] **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

[51] Observer Research Foundation (Saltman, E.), 2022. Identifying and Removing Terrorist Content Online: Cross-Platform Solutions. [accessed 4 July 2023].

[52] Tech Against Terrorism, 2021.

[53] Texas National Security Review (Fishman, B.), 2019. Crossroads: Counter-terrorism and the Internet. [accessed 3 July 2023]

[54] Tech Against Terrorism, 2021. Terrorist use of E2EE: State of play, misconceptions and mitigation strategies. [accessed 3 July 2023].

[55] 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].

[56] Tech Against Terrorism, 2022. Terrorist content analytics platform: year one: 1 December 2020 – 30 November 2023. [accessed 3 July 2023].

[57] Tech Against Terrorism, 2023. Patterns of Online Terrorist Exploitation. [accessed 3 July 2023].

[58] United Nations Office on Drugs & Crime report, 2021.

state or a prominent political figure, rewarding virtual successes, and may be offered in multiple languages to appeal to a broad audience.[59]

Marketplaces and listings services, fundraising services, and payment services

6B.29 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.[60] For more information on this, see Transactions and offers below.

Risk factors: User base

User base size

6B.30 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.[61] 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.

6B.31 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

6B.32 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.

6B.33 Data suggests that user base characteristics including **race and ethnicity, religion,** and **gender** could lead to an increased risk of harm to individuals.

6B.34 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.[62] 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.[63]

6B.35 Data from the Online Experiences Tracker shows that males (4%) are more likely than females (2%) to have experienced or encountered content that encouraged extremism, radicalisation or terrorism in the past four weeks. Those from minority ethnic groups (4%) are more likely than the average (3%) to encounter this content and those aged 18-34 (4%) are also more likely than average to do so (3%).

[59] United Nations Office on Drugs & Crime report, 2012.
[60] United Nations Office on Drugs & Crime report, 2012.
[61] United Nations Office on Drugs & Crime report, 2012.
[62] Foreign Policy Magazine (Ware, J. and Clarke, C.P.), 2022. How Far-Right Terrorists Choose Their Enemies3. [accessed 3 July 2023]
[63] Ofcom, 2022. The Buffalo attack: Implications for online safety. [accessed 29 June 2023].

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles and fake user profiles

6B.36 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.

6B.37 Research indicates that anonymous user profiles can empower potential perpetrators by emboldening them to speak and act more radically.[64]

6B.38 Users can evade disciplinary action taken by services by creating a different user profile with an altered username. A report from the Institute for Strategic Dialogue (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.[65]

User networking

User connections and user tagging

6B.39 The way in which users can find or encounter like-minded individuals on services is key to 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.

6B.40 Research by the Global Network on Extremism and Technology (GNET), an academic and research initiative, points to pro-ISIS[66] accounts on mainstream services which have used the "*user tagging function on posts to tag or link to nearly 100 similar accounts".*[67] 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*".[68]

6B.41 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 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*

[64] 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].

[65] ISD is an organisation dedicated to reversing rising extremism worldwide. Source: Institute for Strategic Dialogue (O'Connor, C.) 2021. Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. [accessed 4 July 2023].

[66] 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].

[67] 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].

[68] GNET, 2022.

even exceed the audiences of many official ISIS channels found on encrypted alt-tech[69] *platforms"*.[70]

User communication

Livestreaming

6B.42 There are many examples of livestreaming being used to promote terrorism. For instance, in 2013, an attack in Kenya was live-tweeted;[71] the attack in Buffalo, New York was livestreamed (and versions of the footage were disseminated on multiple online services);[72] and an attack in Christchurch, New Zealand was streamed live.[73] [74] 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.[75] The use of livestreaming remains a persistent feature of far-right lone attackers, many of whom directly reference and copy aspects of previous attacks.[76] The risks associated with such content are linked to the ability for the content to go viral and motivate others to carry out such attacks in a similar manner.[77] This is mostly linked to the ability for the content to be recorded, the details of which can be found at paragraph 6B.62.

Live audio

6B.43 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 behaviours leading to these gaming-related online "*spaces providing extremists with the opportunity to broadcast their messages widely, and relatively undisturbed"*.[78]

[69] "*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].

[70] 2022 GNET article.

[71] *"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, New York (2020). Chapter 7, Page 189.

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

[73] *"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: Wbur, (Brooks, A and Matromarino, J.P.), 2022. Extremists exploit gaming networks and social media to recruit and radicalize. [accessed 19 September 2023].

[74] 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].

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

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

[77] Bryden, M., Bahra, P., Cruickshank, P., Macklin, G., Cook, J., Vale, G and Simcox, R., 2019. CTCSENTINEL. [accessed 26 July 2023]; Kupper, J., Chrsitensen, 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. [accessed 4 July 2023].

[78] United Nations Office of Counter-Terrorism, 2022. Examining the Intersection Between Gaming and Violent Extremism. [accessed 4 July 2023].

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

6B.44 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.

6B.45 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.[79]

6B.46 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 above features are preferred by terrorist actors.[80]

6B.47 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.[81]

Posting content (text, images, videos)

6B.48 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.

6B.49 For example, an article describes how, on the 20th anniversary of the September 11 attacks, a coalition of alt-jihadist[82] 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*[83] *tested on Discord*".[84] The researchers found that key accounts across services began creating terrorism content using popular internet memes.[85]

6B.50 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 hazardous materials; and how to plan and execute terrorist attacks".[86] The services make it easy for material to be shared among a large group of people, and can also help build a sense of

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

[80] Tech Against Terrorism, 2021. Terrorist use of E2EE: State of play, misconceptions, and mitigation strategies. [accessed 4 July 2023].

[81] Tech Against Terrorism report, 2021.

[82] *"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].

[83] *"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].

[84] Wired article, 2021.

[85] Wired article, 2021.

[86] United Nations Office on Drugs & Crime, 2012. The use of the Internet for terrorist purposes. [accessed 4 July 2023].

community among individuals in different locations and with different backgrounds, “encouraging the creation of networks for the exchange of instructional and tactical material”.[87]

Commenting on content

6B.51 The ability to leave comments on a post is another way in which terrorism content can be promoted and disseminated. In Gaming & Extremism, The Extreme Right on Discord, the authors discuss several instances in which users shared comments about AWD,[88] *“with some inquiring about how to find AWD’s website and others indicating that they would like to join the group”*.[89]

6B.52 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”*.[90]

6B.53 There is evidence to suggest that terrorist organisations use bots[91] 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 comments pages and spread via bot accounts.[92]

6B.54 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.[93] Examples identified by Bellingcat, an independent investigative journalism group, include memes produced in support of the Christchurch attacker and the Charleston mass shooter.[94]

Transactions and offers

Online payments and crowdfunding

6B.55 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.

6B.56 A UN report says that the way in which terrorists use the internet to raise and collect funds and resources can be categorised into: “*four general categories: direct solicitation, e-*

[87] 2012 United Nations Office on Drugs & Crime.

[88] Atomwaffen Division (AWD), an extreme right-wing group.

[89] ISD (Gallagher, A., O’Connor, C., Vaux, P., Thomas, E., Davey, J.), 2021. Gaming and Extremism, The Extreme Right on Discord. [accessed 4 July 2023].

[90] GNET article, 2022.

[91] 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.

[92] Silva, S., 2020. Islamic State: Giant library of group's online propaganda discovered, *BBC,* 4 September. [accessed 19 September 2023]; ISD, 2020. Click reveals ISD discovery of huge pro-ISIS online cache. 8 September. [accessed 19 September 2023].

[93] GNET article, 2022.

[94] Bellingcat, (R. Evans), 2021. White Boy Summer, Nazi Memes and the Mainstreaming of White Supremacist Violence. [accessed 4 July 2023].

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. Funds transfers are often made by electronic wire transfer, credit card or alternate payment facilities available via services such as PayPal or Skype".[95]

6B.57 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 in fact donations are used to fund acts of terrorism".[96]

6B.58 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 i.e. 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.[97]

Content navigation

Hyperlinking

6B.59 Hyperlinking allows terrorist groups or actors to disseminate information to a large audience.

6B.60 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.[98]

6B.61 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 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.

Content storage and capture

Screen capturing or recording

6B.62 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.

6B.63 U2U services which enable livestreaming have been used to broadcast terrorist attacks in real time, as explained in paragraph 6B.42. Although content may only be live for a short period of time, there is evidence to suggest that the content can be recorded and re-shared

[95] United Nations Office on Drugs & Crime report, 2012.
[96] United Nations Office on Drugs & Crime report, 2012.
[97] GNET (Davis, J.), 2021. Technology and Terrorist Financing. [accessed 4 July 2023].
[98] Observer Research Foundation (Saltman, E.), 2022. Identifying and Removing Terrorist Content Online: Cross-Platform Solutions. [accessed 4 July 2023].

or forwarded, thereby increasing its virality[99] and the risks of harm to many individuals. This can be done by using the functionality on 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.[100]

6B.64 Terrorist organisations will consider the storage capacity of a service along with its reach, usability, and other security features.[101] 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*".[102]

Content editing

Editing visual media

6B.65 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 of younger people.

6B.66 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'[103] and the digital communities around them, across many services. Approximately a fifth of these pieces of content supported militant groups, including Hamas and jihadist organisations.[104] There is also evidence that steganography, the hiding of messages in images, is in widespread use by terrorist organisations.[105]

Recommender systems

Content recommender systems

6B.67 Although terrorism content is 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 due to a variety of reasons such as a classification error by automated content moderation systems or poor judgement by a human moderator.

6B.68 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 user being exposed to it.

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

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

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

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

[103] "*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 Research Team., 2022. Looking Beyond the Traditional threat: Alt-Jihadism. [Accessed August 23 2023].

[104] Wired article, 2021.

[105] United Nations Office on Drugs and crime report, 2012.

6B.69 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'[106] 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. [107] 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

6B.70 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.

6B.71 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 ensuring that bad actors have no access to advertiser funding"*[108] (Global Alliance for Responsible Media) show how advertisers have a role in protecting individuals against a harm such as terrorism.

6B.72 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 terrorist 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

6B.73 advertisers, and has led to user distrust".[109]

[106] 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].

[107] 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].

[108] World Federation of Advertisers, 2020. Marketing leaders take action on harmful online content. [accessed 4 July 2023].

[109] 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].

Commercial profile

Capacity and maturity

6B.74 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.[110]

6B.75 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 under 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.[111]

6B.76 Different research conducted by Tech Against Terrorism concludes that smaller and newer services are most at risk of exploitation,[112] as terrorists and violent extremists such as ISIS may use them.[113] 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.[114]

6B.77 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.[115]

[110] 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.

[111] Tech Against Terrorism report, 2021.

[112] Tech Against Terrorism, 2019. Analysis: ISIS Use of smaller platforms and the DWeb to share terrorist content – April 2019. [accessed 4 July 2023].

[113] 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.

[114] 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].

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

6C. Child Sexual Exploitation and Abuse (CSEA)

Warning: This chapter contains content that may be upsetting or distressing in relation to child sexual exploitation and abuse.

Introduction

6C.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the CSEA 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').

6C.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.

6C.3 Schedule 6 of the Act sets out a large number of priority offences relating to CSEA. **We recognise that the CSEA offences are complex and that a particular instance of exploitation or abuse may involve the commission of a number of these specific offences. However,** for the purpose of presenting our analysis, we have split our assessment of the relevant risks into two broad categories: grooming, and child sexual abuse material (CSAM).

6C.4 Some specific types of CSEA can feature both elements of grooming and elements of CSAM, and so are difficult to cover comprehensively in only one of those sections. This is particularly true for offences with an element of sexual exploitation, self-generated indecent images (SGII) and child-on-child sexual exploitation and abuse.

6C.5 Ofcom recognises that these types of CSEA can be closely related and can happen either in isolation, in parallel or sequentially.[116] Offences[117] related to child-on-child sexual exploitation and abuse cut across the different types of CSEA. Our analysis of online grooming and CSAM recognises that the perpetrators of CSEA can themselves be children, particularly in the non-consensual sharing of self-generated indecent imagery (SGII).[118]

[116] 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) (Volume 5, Chapter 26).

[117] 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).

[118] Self-generated indecent imagery (SGII) is CSAM produced by a child, in which a perpetrator does not appear to be present.

6C.6 Child sexual exploitation and abuse (CSEA) involves forcing or enticing a child or young person to take part in sexual activities, not necessarily involving a high level of violence, 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 be used to facilitate offline abuse.[119] Ofcom recognises that the experience of each victim and survivor is unique. We describe the impact of these experiences to help services understand the risk of the harm. We draw attention to the complex effects and trauma that child sexual exploitation and abuse has on victims and survivors. The observations are drawn from research with individuals with lived experiences of the harm, and not an attempt to speak on behalf of those who have experienced this harm.

6C.7 **Although we cover CSEA offences throughout this chapter, we acknowledge that CSEA can manifest in different and complex ways. For services gathering a more complete picture of the risks of harm to children from illegal content, please also refer to the unlawful immigration and human trafficking chapter (relating specifically to the sexual exploitation of children**[120] [121] **- chapter 6J: Unlawful immigration and human trafficking)], chapter 6M: Intimate image abuse and chapter 6S: Cyberflashing. Children may be affected by illegal content described in other chapters as well.**

Relevant offences

6C.8 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. In regard to CSEA, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 6 of the Act.

6C.9 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of the offences listed below.

6C.10 For more details on the offences listed below, and how services can assess whether content amounts to illegal content, please refer to the **Illegal Content Judgements Guidance** (see Volume 5, Chapter 26).

[119] HM Government, 2018. Working Together to Safeguard Children. [accessed 22 September 2023].

[120] 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 **chapter 6J: Unlawful immigration and human trafficking**. Source: Department for Education, 2017. Child sexual exploitation. [accessed 22 September 2023].

[121] 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).

Grooming

6C.11 We consider the following grooming offences:

- causing or inciting a child to engage in sexual activity[122]
- engaging in sexual activity in the presence of a child[123]
- causing a child to watch a sexual act [124] or to look at a sexual image[125]
- arranging or facilitating commission of a child sex offence[126]
- meeting a child following sexual grooming[127] or following certain preliminary contact[128]
- sexual communication with a child[129] or communicating indecently with a child[130]

6C.12 The commission of these grooming offences typically involves an abuser communicating with a child, to facilitate online or offline child sexual abuse. This could involve, for example, sexual communication with a child. The facilitation of these offences may involve an abuser making contact and regularly communicating with a child to develop a relationship before the abuser introduces a sexual element into the conversation.

Child sexual abuse material (CSAM)

6C.13 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[131]
- possession of prohibited images of a child[132]
- possession of a paedophile manual[133]
- taking/making, distribution, possession and publication of indecent photograph or pseudo-photograph of a child[134]
- offences relating to taking/making, distribution, possession, publication etc of indecent photographs etc. of children in Scotland[135]
- offences relating to taking, distribution, possession, publication etc of indecent photographs of children in Northern Ireland[136]

[122] 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)), and section 31 of the Sexual Offences (Scotland) Act 2009. Section 54 Of the Sexual Offences (Scotland) Act 2009.

[123] 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)).

[124] 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)).

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

[126] 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)).

[127] 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)).

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

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

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

[131] 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.

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

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

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

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

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

6C.14 CSAM refers to images or videos which involve penetrative or non-penetrative sexual acts with children, or any other indecent or prohibited imagery of children.[137] This includes 'pseudo-photographs'[138] that have been made entirely on a computer. Perpetrators of CSAM may be involved in the creation, uploading, sharing and distribution of CSAM. Users can also facilitate the creation, uploading, sharing or distribution of CSAM by signalling that they are in possession of it, and sharing links or advising other users where to find it.

How CSEA offences manifest online

6C.15 This section is an overview which looks at how CSEA offences manifest online, and how individuals may be at risks of harm.

6C.16 CSEA is a severe and rapidly evolving threat. It is difficult to estimate the scale of online and offline CSEA for several reasons, including the lack of reporting, the complexity of the reporting process for victims, and the hidden nature of the offending. The Crime Survey for England and Wales (CSEW) estimated in 2019 that 7.5% of adults aged 18 to 74 had experienced sexual abuse before the age of 16 – approximately 3.1 million people.[139] The Independent Inquiry into Child Sexual Abuse (IICSA) reported that research indicates that one in six girls and one in 20 boys are sexually abused before the age of 16.[140]

6C.17 In 2021/22 over 103,000 child sexual abuse offences were recorded by police in England and Wales.[141] The National Crime Agency (NCA) estimates that there are between 680,000 and 830,000 UK based adult offenders who pose varying degrees of risk to children, equivalent to 1.3% to 1.6% of the UK adult population.[142]

6C.18 Reported CSEA figures are unlikely to represent the true scale of online child sexual exploitation and abuse. IICSA reported that CSEA is significantly under-reported.[143] The CSEW found that 76% of adults who experienced rape or assault by penetration before the age of 16 did not tell anyone at the time, and only 7% of victims and survivors told the police at the time.[144] Similarly, 67% of victims and survivors who took part in IICSA's Truth Project did not tell anyone that they were being sexually abused at the time.[145]

[137] This is explained in greater detail in the relevant section below.

[138] A pseudo-photograph is an image made by computer-graphics or otherwise which appears to be a photograph. Source: Home Office, 2023. Indecent Images of Children: guidance for young people. [accessed 22 September 2023].

[139] Office for National Statistics, 2020. Child sexual abuse in England and Wales: year ending March 2019. [accessed 31 July 2023]

[140] Independent Inquiry into Child Sexual Abuse, 2022, The Report of the Independent Inquiry into Child Sexual Abuse [accessed 2 October 2023]

[141] Centre of expertise on child sexual abuse (Karsna, K. and Bromley, P.), 2023. Child sexual abuse in 2021/22: Trends in Official data. [accessed 31 July 2023]

[142] National Crime Agency, 2023. National Strategic Assessment 2023 for Serious and Organised Crime. [accessed: 2 October 2023]

[143] The Independent Inquiry into Child Sexual Abuse, 2022. The Report of the Independent Inquiry into Child Sexual Abuse, [accessed: 02 October 2023]

[144] Office for National Statistics, 2020. Child sexual abuse in England and Wales: year ending March 2019. [accessed 31 July 2023]

[145] 5,862 victims and survivors participated in the Truth Project. The Independent Inquiry into Child Sexual Abuse, 2022. The Report of the Independent Inquiry into Child Sexual Abuse [accessed: 2 October 2023]

6C.19 In recent years, factors relating to the Covid-19 lockdown,[146] including more time spent online,[147] and attempts to reduce isolation by connecting with people virtually, are likely to have contributed to the increased prevalence of online CSEA offending.

6C.20 The evidence indicates that the scale of online CSEA is increasing year on year. In the US, the National Centre for Missing & Exploited Children (NCMEC)[148] CyberTipline received 32,059,029 reports from the public and electronic service providers of suspected child sexual exploitation in 2022, a 9% increase on the previous year and a 47% increase since 2020.[149] NCMEC also saw an 82% increase in incidences categorised as 'online enticement'[150] in 2022, compared to the previous year.[151] There is evidence that incidences of grooming online are increasing. The National Society for the Prevention of Cruelty to Children (NSPCC) found that 6,156 'sexual communication with a child' offences were recorded by the police in England and Wales between April 2021 and March 2022;[152] an average of about 120 offences per week, representing an 80% increase over four years.[153] [154] In addition, Ofcom research found that one in ten children have been asked to share naked or half-dressed photos or videos.[155]

6C.21 The presence of CSAM online is also increasing. There have been year-on-year increases in the number of URLs which contain CSAM reported to the Internet Watch Foundation (IWF), with a 707% increase between 2014 and 2021.[156] 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.[157]

6C.22 Self-generated indecent imagery (SGII) is CSAM produced by a child, in which a perpetrator does not appear to be present. SGII is discussed in greater detail in the CSAM section below. In many cases, a child will have been coerced or directed to produce the indecent imagery (see Grooming section below for further information).[158]

[146] From 2020 – 2021.

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

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

[149] National Centre for Missing & Exploited Children (NCMEC), 2023. CyberTipline 2022 Report. [accessed 31 July 2023].

[150] The term 'enticement' in this context is considered to be close to the term 'grooming' used throughout this chapter.

[151] National Centre for Missing & Exploited Children (NCMEC), 2023. CyberTipline 2022 Report. [accessed 31 July 2023].

[152] Analysis of Freedom of Information requests sent to 41 police forces in England and Wales.

[153] NSPCC, 2022. Online grooming crimes have risen by more than 80% in four years. [accessed 31 July 2023]

[154] Although these figures show an increase in grooming offences there are other factors that have likely influenced these figures, including the changing landscape of relevant legislation, namely the Sexual Offences Act 2003 coming into force in 2017, and in turn changes to the national policing response, recording and convictions of associated offences.

[155] Note: the research did not specify if these requests were unwelcome and whether they came from peers, strangers or adults. Ofcom, 2023. *Understanding Online Communications Among Children* Quant Research. This research is published alongside this consultation.

[156] The Police Foundation (Skidmore, M., Aitkenhead, B., Muir, R.), 2022. *Turning the tide against online child sexual abuse*. [accessed 31 July 2023].

[157] Internet Watch Foundation, 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].

[158] SGII is explored in more detail in paragraphs 6C.40, 6C.49 and 6C.113.

Risks of harm to individuals presented by online CSEA offences

6C.23 The threat presented by online CSEA is constantly evolving as technology develops and perpetrators identify new ways to use services to engage children and offend. For example, offenders are sharing information on how to exploit weaknesses in the implementation of extended reality technologies and artificial intelligence services.

6C.24 Online CSEA can take many different forms. Perpetrators use different functionalities and services depending on the type of offences they are committing, and will move children across services to commit their offences. For example, a perpetrator may first contact a child on one site and then move them to another to share images using encrypted services. The CSAM and Grooming sections below provide more detail on how these offences take place online.

6C.25 Much of the knowledge regarding victims' and survivors' experiences of CSEA is retrospective. Victims and survivors of child sexual exploitation and abuse often take longer than five years (with an average of 17 years) to disclose the abuse.[159] The Crime Survey for England and Wales estimated that of those who had experienced sexual abuse before the age of 16, for 48% of the victims and survivors the sexual abuse had started, or occurred, before the age of 11.[160] [161]

6C.26 Most of the research has highlighted that CSEA tends to be targeted towards girls;[162] girls are twice as likely as boys to receive sexual solicitation requests.[163] Reporting of male experiences of abuse is generally understood to be lower because of the wider social stigma associated with male victims of CSEA, but there has been a recent increase in reports relating to financially motivated sexual extortion.[164]

6C.27 Perpetrators may also seek out children who display certain vulnerabilities online; these can include having mental health difficulties or feeling lonely, or being in care.[165] [166] There is also emerging research on LGBTQ+ children's vulnerability to online sexual solicitation; they have been found to be more likely than their heterosexual peers to start romantic relationships

[159] 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].

[160] Office for National Statistics, 2020. Child sexual abuse in England and Wales: year ending March 2019. [accessed 31 July 2023].

[161] The Independent Inquiry into Child Sexual Abuse found that 79% of the participants in the Truth Project were aged 11 or under when they were first sexually abused. Source: Independent Inquiry Child Sexual Abuse, 2022. The Report of the Independent Inquiry into Child Sexual Abuse. [accessed 31 July 2023].

[162] Office for National Statistics, 2020. *Child sexual abuse in England and Wales: year ending March 2019.* [accessed 31 July 2023]; Internet Watch Foundation, 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].

[163] Finkelhor, D., Mitchell, K. J., and Wolak, J., 2000 cited in 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*, vol.18, pp.135-146.

[164] FBI National Press Office, 2022. FBI and Partners Issue National Public Safety Alert on Financial Sextortion Schemes. [accessed 22 September 2023]; Canadian Centre for Child Protection, 2022. An analysis of financial sextortion victim posts published on R/Sextortion. [accessed 2 August 2023].

[165] 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.

[166] Centre of expertise on child sexual abuse (Karsna, K. and Bromley, P.), 2023. Child sexual abuse in 2021/22: Trends in Official data. [accessed 31 July 2023].

online.[167] Evidence suggests that LGBTQ+ children may face an increased risk of harm online.[168]

6C.28 CSEA can have a profound and long-lasting impact. Victims and survivors often describe how being sexually abused as a child has affected different aspects, and stages, of their lives. In an analysis of victims' and survivors' experiences of child sexual abuse, 88% reported that it had had a negative impact on their mental health.[169] In the same study, almost a fifth of victims and survivors said they had attempted suicide. Research has also highlighted the association between CSEA and post-traumatic stress disorder, depression, anxiety, and personality disorders.[170] [171] Some victims and survivors describe CSEA as having had a negative impact on their education and employment prospects. It is also common for victims and survivors to speak about difficulties in their relationships, often resulting from a lack of trust caused by being sexually abused, or difficulties with intimacy or affection.

6C.29 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.

6C.30 Child sexual abuse victims and survivors are known to experience re-traumatisation and continued re-victimisation as a result of knowing about, or inadvertently seeing, images of their own abuse circulating online.[172]

6C.31 Perpetrators of grooming offences can come from all age groups and demographics. Research in this area is not conclusive, with some research suggesting that some perpetrators may have experienced problems with intimacy, loneliness or self-esteem, or suffer from depression or other mental health issues.[173] Indeed, while some research has stated that perpetrators of online child sexual exploitation and abuse are more likely to be unemployed,[174] other research has found the opposite.[175] It is generally agreed in the literature that perpetrators will have a sexual interest in children, although this may be a latent interest, which is then triggered by the availability of CSAM and access to children

[167] 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.* [accessed 7 August 2023].

[168] Thorn, 2023. LGBTQ+ Youth Perspectives: How LGBTQ+ Youth are Navigating Exploration and Risks of Sexual Exploitation Online. Findings from 2022 qualitative and quantitative research among 13-20-year-olds. [accessed 16 August 2023].

[169] 5,862 victims and survivors participated in the Truth Project. Independent Inquiry Child Sexual Abuse 2022. The Report of the Independent Inquiry into Child Sexual Abuse. [accessed 22 September 2023].

[170] 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.

[171] Maniglio, R., 2009. The Impact of Child sexual abuse on health: a systematic review of reviews. *Clinical Psychology Review*. 29(7), pp.647 – 657.

[172] A campaign by the Canadian Centre for Child Protection (C3P) in 2021 highlighted that some victims and survivors actively search for their own CSAM online in an effort to get it removed. Source: Canadian Centre of Child Protection. 2021. Happy 15th Birthday, Twitter.[accessed 7 August 2023].

[173] 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].

[174] Babchishin, K.M., Karl Hanson, R., Hermann C. A., 2010. The Characteristics of Online Sex Offenders: A Meta-Analysis. *Sexual Abuse*. 23(1) pp.92-123.

[175] Chopin, J., Paquette, S. and Fortin, F., 2022. Geeks and Newbies: Investigating the Criminal Expertise of Online Sex Offenders. *CrimRxiv*.

online.[176] Some perpetrators may not have a sexual interest in children themselves but may have other motivations, including financial gain.[177]

6C.32 It is challenging to estimate the economic and social cost of CSEA offences; however, we present an estimate by the UK government to give an indication of the potential scale. The UK government estimated the economic and social cost of contact child sexual abuse in England and Wales. Accounting for inflation, this is approximately £101,700 per victim in 2022 prices, although it is acknowledged that this is likely to be an underestimate.[178]

[176] 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 behavior*. 44.

[177] 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. Source: Internet Watch Foundation, 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].

[178] The original study estimated that the cost per victim was £89,240 (in 2018/19 prices). This was a deliberatively conservative estimate of the cost of CSA. For example, the estimate only reflects the cost of non-fatal CSA (i.e. does not apply to fatal CSA) and does not account for long-term mental health impacts. Source: Home Office, 2021. The economic and social cost of contact child sexual abuse. [accessed 22 September 2023].

Grooming

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

This chapter looks at the evidence surrounding online grooming for the purposes of conducting child sexual abuse. Grooming is a complex process that often does not follow a clear offending pathway and can be a protracted process over weeks or months, or occur rapidly following initial contact. There are many stages in the process of a perpetrator grooming a child. Grooming is not limited to sexual communications, although it can involve coercing or manipulating children into multiple sexual acts.

Grooming will affect each victim and survivor differently, but the effects are significant and can often last a lifetime. These include negative psychological impacts, including self-harm, loss of confidence and aggression, increased feelings of self-blame and lack of personal trust.

Service type risk factors:

Perpetrators can groom children in a multitude of different ways, and there are many different services where grooming is conducted. **Social media services, gaming services** and **messaging services** are commonly used in propagating this offence. They have therefore been included in the risk profiles. Other service types also accessed by children (**video-sharing services, discussion forums and chat rooms**) are also relevant to cases of grooming.

Groomers can 'platform-hop' across services to take advantage of different functionalities enabling them to meet children, engage in grooming conversations and avoid detection.

User base risk factors:

For grooming offences, a high-risk factor in terms of user base characteristics is the **age of users.** Groomers who want to contact child users will be drawn to services that children access. It is therefore reasonable to assume that, whilst grooming can take place on services of all sizes, services with larger numbers of child users than other services, or a particularly high proportion of child users, are in some cases likely to present more of an incentive to perpetrators in the initial stages of grooming, such as identifying and initiating contact with children. This characteristic is likely to be less important once contact has been established. In the later stages of grooming a service's perceived lack of detection technology may be a more defining risk characteristic. However, the ability for a child to access a service will remain a constant factor, given the nature of the offence.

Perpetrators will assess a child's personal characteristics to identify traits which could make them vulnerable to online grooming, such as low self-esteem or a lack of supervision. Other factors may also make children more vulnerable to online grooming, including identifying as **LGBTQ +** and/or having a **disability**. **Gender** appears to be a risk factor in grooming offences; the evidence suggests that girls are more likely to be groomed than boys. But it is acknowledged that there is significant under-reporting by male victims.

Functionalities and recommender systems risk factors:

Functionalities that allow abusers to identify and make contact with children are risk factors in the facilitation of grooming offences. **User profiles**, and the information that is presented on them, can be used by perpetrators to identify and target victims and survivors, thereby starting the grooming process. The ability to create **fake user profiles** allows perpetrators to misrepresent themselves to victims and survivors by displaying a false age, name and location. **User connections** allow perpetrators to establish contact with child users who have been identified and begin communicating. The sense of trust that mutual connections can create may also be exploited by perpetrators. Perpetrators can often use network recommender systems in the form of publicly displayed connections list of children to infiltrate groups of children at speed and utilise this as a tool to blackmail and coerce children to further the sexual abuse. **User groups** can be used to target children, particularly groups for adolescents that discuss sexual themes. Due to their role in propagating grooming, user profiles, fake user profiles, user connections and user groups have been included in the risk profiles.

Direct messaging can allow perpetrators to establish contact with children and potentially develop relationships through frequent communication that often takes place away from public view. **Encrypted messaging** makes detecting perpetrators' contact with children challenging. Messaging functionalities of direct messaging and encrypted messaging are also included in the risk profiles.

Network recommender systems can facilitate grooming by suggesting adult users to children. Child users, and groups often associated with children, can also be suggested to perpetrators, based on their past activity and connections. For these reasons, network recommender systems have also been included in the risk profiles.

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 with children, as well as exchange contact details. Livestreaming and commenting on content have therefore also been included in the risk profiles.

Other functionalities also play a role in this offence. For example, **sending images through messaging functionalities** can be used to persuade children to share SGII. **Visual media editing** functionalities can also be used to disguise the identity of perpetrators when they are contacting a child.

How grooming offences manifest online

6C.33 This section is an overview which looks at how the specified grooming offences manifest online, and how users may be at risk of harm.

6C.34 Online grooming for child sexual abuse is the method of contacting children and developing a relationship, whether through flattery, emotional connection, sexualisation, bribery, blackmail or coercion, for the purposes of conducting child sexual abuse. Typically, the objective of online grooming is the generation of child sexual abuse material (CSAM) and contact sexual abuse of children. Contact sexual abuse of a child can occur in person or can involve the perpetrator remotely forcing the victim to abuse other children or to engage in sexual acts, including penetrative acts. Individuals involved are often incited to share imagery of the abuse with the perpetrator as 'first-generation' CSAM.[179]

6C.35 Typically, grooming for sexual abuse will first involve identifying a child. Then, a perpetrator will seek to make contact and communicate with the child. Although an offence may not take place in the identification of the child, it is nevertheless a crucial step towards the commission of grooming offences. The evidence presented here concerns both these stages.

6C.36 The research highlights that grooming is significantly under-reported by victims for many reasons including shame, fear, and the lack of recognition of the crime by some children who experience it.[180] While it is not possible to accurately determine the scale of online grooming, it is understood that '*the scale of online grooming is of real and significant concern*.'[181]

6C.37 It is therefore difficult to capture the exact scale of online grooming due to complexities in how it presents, and how it is experienced and identified.

6C.38 Reports from law enforcement in the UK show that in 2021/22, police forces recorded 70 different apps and games involved in reported online grooming offences.[182] Since 2017 more than 27,000 'sexual communication with a child' offences have been recorded, with offences increasing by 80% between April 2017 and March 2021.[183] A US study among undergraduates found that 17% of participants had experienced sexual solicitation as youths from adults they had chatted with online and 23% recalled a long intimate conversation with an adult stranger which could be indicative of online grooming.[184]

6C.39 The prevalence of unsolicited sexual messages that children receive from unknown users is also an indicator of online grooming. Ofcom research looking at young people's communication online has found that 30% of children aged 11-18[185] years old claim to have ever received unwanted friend or follow requests, 13% of 11-18-year-olds say they have

[179] 'First-generation' or 'novel' CSAM refers to material that is newly generated and which has not been previously shared or re-shared, explored further in the CSAM section below.

[180] 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.

[181] Independent Inquiry into Child Sexual Abuse, 2020. The Internet: Investigation Report. [accessed 31 August 2023].

[182] NSPCC, 2022 Online grooming crimes have risen by more than 80% in four years. [accessed 22 September 2023].

[183] NSPCC, 2022 Online grooming crimes have risen by more than 80% in four years. [accessed 22 September 2023].

[184] 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.

[185] 18-year-olds in this study were asked to reflect back on their experiences before turning 18.

ever received pictures or videos of naked or half-dressed people when communicating online, and 10% have ever been asked to share an intimate picture or video of themselves.[186] In addition, 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.

6C.40 The prevalence of self-generated indecent imagery (SGII) is also indicative of the scale of grooming, as some SGII is coerced from children via grooming. The IWF identified that CSAM created by children accounted for almost four in every five (78%) reports it dealt with in 2022. Many of these images will be produced as a direct result of a child being groomed, however the exact proportion of SGII resulting from coercion and grooming is not known.[187]

6C.41 It is difficult to define the exact type of service where grooming offences take place, and it can occur on both large and small services. Grooming often occurs in stages which take place on multiple different services.[188] The risk to children from grooming presents itself at multiple points along the grooming journey, from identification through to direct communication and enticement or coercion.

6C.42 There are a few key enablers of grooming offences. These include the presence of children on a service. This is crucial, particularly for the first stage, when perpetrators identify and initiate contact with a child online. When initiating contact, perpetrators may seek out services with a high proportion of child users. They may also seek out services that encourage new connections, to quickly initiate contact with a child or children.

6C.43 Another key component of grooming is sexual communication. Perpetrators will seek out spaces that enable sexual communication, allowing the grooming process to proceed. Service-hopping is common in grooming offences, as it allows perpetrators to isolate the child and to access functionalities, such as image-sharing, to facilitate the sexual abuse.[189] According to Ofcom research, 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.[190] [191]

6C.44 Many perpetrators seek to move children to other online spaces, such as private messaging services, to continue the grooming process, so that they can make use of functionalities such as private chat or end-to-end encryption and avoid content moderation and detection. The different types of risk linked to grooming are therefore likely to vary depending on the functionalities of a service, making it important that services are aware at what stages in the grooming process their service may be exploited, so that they can deploy proportionate mitigations.

[186] Ofcom, 2023. *Understanding Online Communications Among Children* Quant Research. This research is published alongside this consultation.

[187] Internet Watch Foundation, 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].

[188] Movement through the stages is often driven by the preparator's motives. Perpetrators' objectives range from in-person sexual activity with a child, to coercing children to sexually abuse others, defined as 'child sexual abuse by proxy', and often involving the creation of first-generation CSAM which the perpetrator can then share online.

[189] 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.

[190] This 'uncomfortable experience' may not be grooming *per se*, but may be intimate image sharing, rude/abusive messages or being asked for personal information.

[191] Ofcom, 2023. *Understanding Online Communications Among Children* Quant Research. This research is published alongside this consultation.

Risks of harm to individuals presented by online grooming offences

6C.45 Grooming is a behavioural offence that can take many different forms online and does not follow any one defined pattern. 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.

6C.46 Existing literature has highlighted key offender patterns that occur in most grooming cases. These generally include phases of friendship forming, trust development, risk assessment, exclusivity, conversation related to sex, and a conclusion, namely the online and/or offline sexual abuse of a child.[192] These stages are not fixed, and often occur interchangeably and/or extremely rapidly. They merely provide an indication of the key points in an offending journey where possible friction points might be applied to deter offenders.

6C.47 Perpetrators, and those seeking to sexually abuse children online, may be known or unknown to the child. Perpetrators seeking to groom children deploy multiple conversational styles and tools that help them to sexually abuse children online. This can include flattery, trust-building, threats, sexualisation and bribery.[193] In some cases, perpetrators impersonate other young people or create fake online user profiles in order to build relationships with children and obscure their identity, using flattery or common interests. In other cases, perpetrators can be truthful about who they are throughout the grooming process. Perpetrators may also try to blackmail the child, which can make it difficult for children to end contact with perpetrators of grooming or other CSEA offences online.[194]

6C.48 Some grooming perpetrators use online services to target large numbers of children, 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.[195] 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.[196] Other techniques used by perpetrators include focusing on relationship-building, such as 'the boyfriend model'[197] where the course of the interaction can last for days or even years.

[192] 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.

[193] 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.

[194] 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.

[195] 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]

[196] 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.]*

[197] Barnardo's, 2017. Working with children who are victims or at risk of sexual exploitation: Barnardo's model of practice. [accessed 10 August 2023].

6C.49 There is a growing trend of SGII being used as a method to blackmail children into sending money. In these cases, a perpetrator will coerce children into producing SGII, often by first sending them sexual images. The perpetrator will then demand a form of payment (usually money or vouchers) or 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.[198] This is dependent upon the perpetrators' motivations for committing grooming offences, typically financially or sexually driven objectives. Reports received by NCMEC in early 2022 found that in 79% of cases involving sexual extortion[199] the perpetrators were seeking money.[200] In 2022, the US received 7,000 reports related to online sexually-coerced financial extortion, with 3,000 victims and more than a dozen related suicides.[201] In the UK, there is also evidence to suggest that many cases of online child sexual abuse involve sexually-coerced financial extortion, particularly involving boys.[202]

6C.50 Grooming will affect each victim and survivor differently, but the effects are significant and can last a lifetime. Qualitative studies based on interviews with individuals who have experienced online grooming highlight the negative psychological impacts, including self-harm, loss of confidence and aggression, increased feelings of self-blame and lack of personal trust.[203] Perpetrators often use tactics such as manipulation and coercion, including inciting victims into sending sexual material, which can increase feelings of self-blame. This often increases victims' and survivors' reluctance to disclose instances of online grooming.[204]

6C.51 Other research has highlighted how a shift occurs in people's online behaviour after the experience; rather than viewing the internet as a place of opportunities, their attitudes shift to a more negative view.[205] Typically, the objective of online grooming is contact sexual abuse of children and/or the creation of CSAM, for which grooming is a facilitator. The impact of CSAM, through SGII or other means, on an individual is discussed in more detail in the CSAM section below. The cycle of abuse can also be escalated, with the perpetrator inciting the child to include other children, objects or animals in the sexual abuse, thereby intensifying the child's complex feelings of shame and culpability.[206]

[198] Several recent cases have received high-profile media attention both in the USA and the UK. Source: CNN, 2023. 'He lost his son to suicide after a 'sextortion' scam. Now this lawmaker is fighting to save other.' [accessed 10 August 2023]; Chad, 2023. 'Police warning after growing number of Nottinghamshire 'sextortion' cases'. [accessed 10 August 2023]

[199] We use the term 'sexually-coerced financial extortion'. This term has previously been used to describe behaviour without a financial incentive, but for the Register we use it to refer to the use of sexual images to blackmail children into sending money. Ofcom is aware that the use of sexually-coerced financial extortion does not accurately represent all the factors that are involved in the taking and sharing of images. However, until there is a more appropriate term, we have chosen to align ourselves with the current terminology used by UK law enforcement.

[200] NCMEC, 2022. Sextortion. [accessed 10 August 2023]

[201] FBI National Press Office, 2022. FBI and Partners Issue National Public Safety Alert on Financial Sextortion Schemes. [accessed 22 September 2023];

[202] For example: Gloucestershire Constabulary, 2022. Warning after growing number of sextortion cases in the county. [accessed 22 September 2023].

[203] 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.

[204] 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.

[205] Chiu, J. & Quayle, E., 2022. Understanding online grooming: An interpretative phenomenological analysis of adolescents' offline meetings with adult perpetrators. *Child Abuse & Neglect*, 128.

[206] Independent Inquiry into Child Sexual Abuse, 2020. The Internet: Investigation Report. [accessed 22 September 2023].

Evidence of risk factors on user-to-user services

6C.52 We consider that the risk factors below are liable 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

6C.53 Perpetrators can groom children in many different ways which are often not confined to a single service. They may engage with a child in multiple online spaces, depending on the specific grooming act. For example, when inciting a child to send SGII, a perpetrator may move the conversation with the child to a service that has end-to-end encrypted messaging.

6C.54 Although grooming offences can happen on a number of different services, we found research on the following service types being used to facilitate or commit grooming offences: discussion forums and chatrooms, social media services, private messaging services and online gaming services.

6C.55 Those seeking to groom children often identify and establish contact with children on services that have open channels of communication, such as discussion forums, chat rooms or gaming services, before moving to a service with more privacy, such as a messaging service.[207]

Discussion forums and chat room services

6C.56 The NSPCC notes 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*".[208] This is supported by evidence from Internet Matters, which shows that online groomers may strike up a relationship with a child through discussion forums or online games before asking them to move to another service to talk privately.[209]

6C.57 Research suggests that young people who struggle to form friendships and relationships offline compensate by seeking interactions in chat rooms.[210] 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

6C.58 Our evidence indicates that social media services and video-sharing services are being used for grooming. Reports of online grooming made to the NSPCC mention children being approached on 'social media networks' and 'livestreaming platforms'.[211] Social media services will often be used by perpetrators to target a large number of young users by sending out multiple connection requests.[212] And the evidence shows that livestreaming, which is a common functionality on video-sharing services, enables the commission of grooming offences[213] such as inciting a child into sexual activity.

[207] 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.

[208] NSPCC, 2020. The impact of the coronavirus pandemic on child welfare: online abuse. [accessed 10 August 2023].

[209] Internet.matters.org, n.d. Learn about online grooming. [accessed 10 August 2023].

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

[211] NSPCC,2020. The impact of the coronavirus pandemic on child welfare: online abuse. [accessed 22 September 2023].

[212] Internet.matters.org, n.d. Learn about online grooming. [accessed 10 August 2023].

[213] NSPCC, 2018. Livestreaming and video-chatting: Snapshot 2. [accessed 10 August 2023].

Messaging services and gaming services

6C.59 Our evidence indicates that direct messaging and group messaging functionalities, which are central to most messaging services, are risk factors. A study of online grooming by Thorn in 2022 notes that "*private messaging apps warrant unique consideration for the role they play in meeting people online and how these relationships deepen for minors*".[214] The study notes that 65% of children surveyed[215] reported having an online-only contact invite them to move from a public chat into a private conversation on a different platform. 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.[216] [217]

6C.60 Research from the NSPCC also notes reports of online grooming in which children have been approached on 'instant messaging apps' as well as the "*voice or text chat services built into online multiplayer games*".[218] CSEA professionals also highlight that gaming services can be exploited by those seeking to groom children,[219] by using a service's features to gain contact and trust with them online. For example, 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.[220] This suggests that messaging services and gaming services can be used to facilitate or commit grooming offences.

Risk factors: User base

User base size

6C.61 Both large and small user bases can be risk factors for online grooming. On a large service, a perpetrator may engage in a scatter-gun approach, randomly targeting large numbers of children, with the large user base providing a wider pool of children to approach. In contrast, a service with a smaller user base may allow a perpetrator to more easily identify a victim with specific characteristics or vulnerabilities.

6C.62 The size of a service can also be important at different points in the offending pathway. In these cases, a service with a larger user base will be important for user identification, whereas a smaller service may be more important for user communications, such as messaging and sharing content. Perpetrators will move victims from larger to smaller services depending on their objective.

[214] Thorn, 2022. Online Grooming: Examining risky encounters amid everyday digital socialization. [accessed 10 August 2023].

[215] 445 9–12-year-olds and 755 13- to 17-year-olds in the United States were surveyed.

[216] Thorn, 2022. Online Grooming: Examining risky encounters amid everyday digital socialization. [accessed 10 August 2023].

[217] From this study it is not possible to infer that these were adult-led conversations, nor that grooming was involved.

[218] NSPCC, 2020. The impact of the coronavirus pandemic on child welfare: online abuse. [accessed 22 September 2023].

[219] Interpol, 2020. Threats and Trends Child sexual exploitation and abuse: covid-19 impact. [accessed 25 August 2023];

[220] See for example: 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', *BBC Nfqews*¸30 May. [accessed 2 October 2023].

User base demographics

6C.63 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.

6C.64 The data suggests that user base characteristics including **age, gender, disability, sexual orientation** and media literacy could lead to an increased risks of harm to individuals.

6C.65 Age is an important risk factor in the commission of grooming offences. Groomers who want to contact child users will be drawn to services that children access.[221]

6C.66 There is evidence that gender is a risk factor in the commission of grooming offences; girls are at greater risk of experiencing grooming online. The NSPCC found that girls represented at least 82% of the grooming cases the police dealt with in 2021/22.[222] [223] However, boys are less likely to report sexual abuse, usually due to the perceived social stigma of this type of crime. This may affect the accuracy of the gender-related data for this harm. Further, evidence suggests that a large proportion of the reports of SGII relating sexually-coerced financial extortion involve boys.[224]

6C.67 Ofcom research found that girls aged 16-18 over-indexed for all ten potentially uncomfortable experiences[225] asked about in the online communication research, compared to other ages and genders. For example, 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.[226]

6C.68 Evidence suggests that neurodivergent children, and those with disabilities such as learning difficulties, could be more vulnerable to online grooming. Similarly, Ofcom research on online communications found that those aged 11-18 with limiting or impacting conditions were more likely than those without any such conditions to encounter all ten potentially uncomfortable experiences asked about in the research, including being asked to share intimate images or being sent intimate images.[227]

6C.69 WeProtect[228] found that children who identify as LGBTQ+, along with children with disabilities, were more likely to experience online sexual harms in their childhood.[229] This

[221] 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.

[222] NSPCC response to 2022 Ofcom Call for Evidence: First phase of online safety regulation.

[223] Whittle et al. (2013) also found that girls may be twice as likely to be groomed. Source: 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.

[224] Internet Watch Foundation, 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].

[225] 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

[226] Ofcom, 2023. *Understanding Online Communications Among Children* Quant Research. This research is published alongside this consultation.

[227] Ofcom, 2023. *Understanding Online Communications Among Children* Quant Research. This research is published alongside this consultation.

[228] An independent organisation that brings together governments, the private sector, civil society and intergovernmental organisations working on CSEA online.

[229] WeProtect Global Alliance, 2021. Global Threat Assessment 2021. [accessed 22 September 2023].

may be because these children are more likely to seek relationships online if they feel that they have little opportunity to explore their sexuality in their local community. Ofcom research on online communication found that those aged 11-18 who identified as LGBTQ+ were more likely than those who identified as heterosexual to encounter all ten potentially uncomfortable experiences asked about in the research, including being asked to share intimate images or being sent intimate images.[230]

6C.70 In terms of media literacy, there is evidence suggesting 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. [231] In these cases, the lack of information from a service on how certain functionalities may increase risk to online grooming leads to uninformed decisions (i.e. child users may be more inclined to post personal information without understanding the personal risks that this may have).

Risk factors: Functionalities and recommender systems

User identification

User profiles

6C.71 User profiles, and the information that is presented on them, can be used by perpetrators to identify and target victims and survivors.

6C.72 The information provided in users' profiles facilitates the commission of grooming offences, as it helps perpetrators identify children to target. Perpetrators have spoken of selecting whom to contact online, based on information provided in user profiles, [232] such as profile pictures, name, age and location.

6C.73 The International Centre for Missing & Exploited Children uses a model[233] of online grooming, [234] 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. This can include assessing whether the child's user profile indicates low self-esteem or a lack of supervision. A large part of a perpetrator's risk assessment is likely to draw from the information provided in user profiles.

6C.74 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. The study found that offenders based their decisions 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.[235]

[230] Ofcom, 2023. *Understanding Online Communications Among Children* Quant Research. This research is published alongside this consultation.

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

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

[233] The International Centre for Missing & Exploited Children (Australia) adapted the Winters & Jeglic Sexual Grooming Model for online perpetrators.

[234] International Centre for missing and exploited children 2022. The new "Stranger Danger": Tactics used in the online grooming of children. [accessed 11 August 2023].

[235] 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.

6C.75 Research by 5Rights Foundation found that child-aged avatar accounts received sexual content from adults within hours of being online. One child, aged 12, spoke of relationships on these platforms quickly becoming sexual.[236] This evidence shows how abusers identify child-aged user profiles and target them.

Fake user profiles

6C.76 The ability to create a fake user profile which allows users to present false representations of themselves appears to be widely regarded as a key tool by perpetrators to facilitate the commission of grooming offences. Abusers can create 'fake' user profiles and present themselves as desirable to their target, through the details of their age, name and location.[237]

6C.77 Similarly, 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.

6C.78 Ofcom research found that 15% of 11-18-year-olds claim to have received a friend request from someone pretending to be someone else.[238] This can involve the perpetrator lying about their age. 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 sexually abusing other children.[239]

User networking

User connections

6C.79 The ability for users to connect with each other facilitates grooming, as it allows perpetrators to establish contact with child users and begin communicating. A known methodology is the 'scatter-gun' or 'pyramid approach' where perpetrators use user connection functionalities to try to access large numbers of children. A study by Kloess *et al*. found an example of an offender randomly adding children to initiate contact with them.[240]

6C.80 The user connection functionality, and the visibility of user connections, also appear to facilitate grooming, as these features instil a sense of 'relationship' among mutual connections. For example, young people's social networking predominantly involves 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 [241] and can produce a false sense of security. For instance, to children it may appear that the abuser is known to their social network.[242] User connections can therefore be exploited by abusers to appear as a trusted contact of mutual friends.

[236] 5Rights Foundation response to Ofcom 2022 Call for evidence: First phase of online safety regulation.

[237] 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].

[238] Ofcom, 2023. *Understanding Online Communications Among Children,* Quant research. This research is published alongside this consultation.

[239] BBC News, 2021. David Wilson: Sex offender who posed as girls online jailed for 25 years. 10 February. [accessed 11 August 2023].

[240] 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.

[241] 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].

[242] Hamilton-Giachritsis, C., Hanson, E., Helen, W., Alves-Costa, F., and Beech, A., 2020.

6C.81 As well as being used to identify targets, user connections can be used to coerce children. Perpetrators will 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.[243]

User search

6C.82 Search functions that enable children, or groups containing children, to be found on services are a risk factor in enabling grooming. Research has shown that the ability to target particular cohorts of potential victims and survivors helps facilitate grooming.[244] Search functionalities can enable perpetrators to more easily identify potential victims based on characteristics of interest.

User groups

6C.83 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.[245]

6C.84 Perpetrators create user groups focused on topics that attract young people in order to capture their interest. They can then impersonate young people, potentially for several months, as a way of developing a bond with their potential targets as part of the grooming process. [246] Some perpetrators work together in such groups to enable grooming.

6C.85 There are also online communities popular among adolescents which focus on explicit sexual discussions and obscene language. [247] User groups can help create such communities. Engaging with other users in such an environment may help desensitise child users to sexual solicitations from perpetrators in these communities.

User communications

Livestreaming

6C.86 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. Livestreaming can be used by perpetrators to have sexualised conversations with children and/or incite children to engage in sexual activity in real time. The NSPCC found that sexualised conversations take place when children livestream. It found that of those children who livestreamed, 6% 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.[248]

6C.87 Livestreaming can also be used as a way of facilitating online CSEA, where children are incited to engage in sexualise acts, sometimes including other children, in real time on

[243] Safer Schools, 2023. Protecting Young People from Sextortion. [Accessed 1 September 2023].

[244] 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].

[245] 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.

[246] 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

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

[248] Survey of nearly 40,000 children aged 7 – 16 years old. Source: NSPCC, 2018. *Livestreaming and video-chatting.* [accessed 22 September 2023].

camera. Livestreaming functionalities and the ability to screen-record have also been used by perpetrators to create permanent records of SGII, which can then be used to carry out sexually-coerced financial extortion.[249]

6C.88 Perpetrators are also known to use livestreaming services to direct other adults to conduct the sexual abuse of children in real time. This is often done for a fee, and could occur within a country or transnationally, where perpetrators in one country watch or direct the sexual abuse of a child in another country.[250]

Video calling

6C.89 Video calling can allow children to communicate with individuals who are unknown to them. It can also be used by perpetrators to ask children to engage in sexual activity or to create sexually explicit material while on the call.

6C.90 A study by the NSPCC found that 12% of children had video-called with someone they did not know in person. During these calls, 10% of primary-aged and 11% of secondary-aged children were asked to change or remove their clothes.[251]

Direct messaging, encrypted messaging, and group messaging (text, images, emojis)

6C.91 The ability to communicate on a regular basis is key to perpetrators establishing a grooming relationship with children.[252] Messaging functionalities allow frequent, swift and relatively private exchanges, where perpetrators can develop relationships with children away from public view and parental supervision.[253] The nature of online communication can reduce barriers such as social status and age that exist in face-to-face environments, enabling feelings of intimacy to grow more quickly and bringing more freedom to broach sensitive topics such as sex.[254]

6C.92 The introduction of end-to-end encryption makes it hard to detect offenders' contact with children. Law enforcement agencies highlight the impact that increased prevalence of end-to-end encryption could have on detecting offenders and child safety.[255]

6C.93 There is evidence of children and young people experiencing potentially uncomfortable experiences on direct messages. Ofcom research found that among children and young people (aged 11 to 18) who had experienced having had a potentially uncomfortable online contact experience, messages (with an individual and in a group) was reported as the main method of communication with the other user(s).[256]

6C.94 As with direct messaging, group messaging provides a way for adults to directly engage with children whom they might not know and develop relationships. Ofcom research found that 23% of children had been added to a group chat which included people they did not know

[249] Internet Watch Foundation, 2018. Record numbers of UK men fall victim to sextortion gangs [accessed 16 August 2023].

[250] This is discussed further in paragraph 6C.161.

[251] NSPCC, 2018. Livestreaming and video-chatting. [accessed 22 September 2023].

[252] 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.

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

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

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

[256] Ofcom, 2023. Understanding Online Communications Among Children Quant Research. This research is published alongside this consultation.

well, or at all – this was the second most common of the potentially uncomfortable online contact experiences asked about. [257]

6C.95 In addition to text, the ability to share images and emojis via messaging functionalities can enable relevant offences. Perpetrators can send images to persuade children to share SGII. To normalise their requests for sexual images, and to lower a child's inhibitions, a perpetrator may first share pornographic images with the child. [258] This is followed by asking the child to send the perpetrator similar images of themselves. See para. 6C.40 and 6C.49 on self-generated indecent imagery and the link with grooming offences.

6C.96 The use of emojis can help perpetrators to incite children to engage in sexual activity. A study in 2020 by Lykousas and Patsakis of discussions on LiveMe found that emojis are used to send sexually suggestive messages. [259] Their analysis of over 39 million chat messages, exchanged by more than 1.4 million users in 291,000 live broadcasts over two years, found that emojis were used as a form of communication to request sexually inappropriate and suggestive acts, such as the removal of clothes. The emojis used in these instances were clothing-related emojis, hand gestures and tongue emojis.

Commenting on content

6C.97 Perpetrators might use comments on uploaded content as a means of building rapport with a victim 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. [260]

6C.98 There is also evidence to suggest that livestream comments are used to facilitate grooming offences (see Livestream section above). Hyperlinks, or attempts to exchange contact details, can be shared in comments, with the aim of getting the child to connect with the perpetrator on another service where the grooming process can begin. As outlined in the Livestream section above, the NSPCC found that sexualised conversations take place when children livestream. [261] Transcripts from the IWF also show perpetrators leaving sexualised comments when children livestream. [262]

[257] Ofcom, 2023. Understanding Online Communications Among Children Quant Research. This research is published alongside this consultation.

[258] 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).

[259] Lykousas, N. and Patsakis, C. 2021. Large-scale analysis of grooming in modern social networks. *Expert systems with applications, 176.*

[260] 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].

[261] NSPCC, 2018. Livestreaming and video-chatting [accessed 22 September 2023].

[262] 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].

Posting or sending location information

6C.99 The sharing of a user's current location could 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 the ability to a child to post or send their location, can enable a perpetrator to build up their knowledge base about a child, such as their school or other places they frequent. Perpetrators can then use this information to gain the trust of children; for example, by establishing they have common links to a particular area. They could use this knowledge as a threat to the young person, in order to further their abuse.

Transactions and offers

Accepting online payments

6C.100 Perpetrators may send money to a child or buy them gifts (either virtual or physical) as part of the grooming process. This can be enabled by a service accepting online payments. Perpetrators give gifts in this manner to flatter and 'express their love' to the child user. [263] This element of relationship building facilitates grooming offences.

6C.101 Evidence also suggests that an increasing number of perpetrators are coercing children into sending self-generated indecent imagery (SGII) by offering them money. 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.[264]

Content storage and capture

Screen recording or capturing

6C.102 Screen-recording and screen-capture functionalities can be deployed by perpetrators during video calls or livestreaming to non-consensually capture indecent images.[265] They can then use these images to blackmail the child to generate further CSAM.

Content editing

Editing visual media

6C.103 Perpetrators have been known to pose as young people when contacting children for the purpose of initiating the online grooming process. Perpetrators use video- or image-editing functionalities, such as image filters or video filters, to disguise their real identity when calling a child or sending them photos of themselves. In this way these functionalities aid the facilitation of grooming offences as they allow perpetrators to create a false identity in order to deceive children. Deepfake technology and advances in generative AI can also assist in the creation of images and videos, which may then be uploaded to user-to-user services and

[263] 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.

[264] BBC News, 2021. Abdul Elahi: Sexual blackmailer jailed for 32 years.[accessed 16 August 2023].

[265] 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.

enable approaches to children to seem more genuine, increasing the chance of a response.[266]

Recommender systems

Network recommender systems

6C.104 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% of children aged 11-18 said that they had added contacts via friend suggestions/ Quick adds/ Connections requests or Discover functions.[267] It is possible that 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.

6C.105 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.

Risk factors: Business models and commercial profiles

Revenue models

6C.106 No evidence was found suggesting that revenue models are a risk factor in the facilitation and commission of these offences.

[266] 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.

[267] Ofcom, 2023. *Understanding Online Communications Among Children.* Quant research. This research is published alongside this consultation.

Child Sexual Abuse Material (CSAM)

Summary analysis for child sexual abuse material: how harms manifest online, and risk factors

CSAM can have a profound and long-lasting impact on children who are sexually abused, and can have an impact on the wellbeing of adults and children who unintentionally view this material.

In addition to the abuse itself, the existence, sharing and viewing of images and videos of that abuse can be a continuing source of trauma for victims and survivors of child sexual exploitation and abuse. Victims and survivors can be re-victimised and also affected by heightened sensitivity to photos and cameras.

Unintentional viewing of CSAM by adults is likely to cause considerable distress and can also lead to difficulties in stopping the viewing of such material, becoming a regular viewer of CSAM. Regular viewing of CSAM can lead to perpetrators making contact with children to perpetrate sexual abuse, both online and offline.

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); this requires a complex understanding of context and is explored further in this chapter.

Service type risk factors:

Any service can be used to distribute CSAM. Services that have the capacity to share images or videos, post text or share hyperlinks pose particular risks.

File-storage and file-sharing services, in particular those that allow users to upload and share images through links, are considered particularly risky, facilitating the storage of large, curated collections of CSAM. For their role in propagating this offence, file-storage and file-sharing services have been included in the risk profiles.

Evidence indicates that other types of service also increase the risk of CSAM offences. These include **social media services**, **discussion forums and chat rooms, messaging services** and **adult services**, and have therefore been included in the risk profiles. **Video-sharing services** are also a risk factor.

User base risk factors:

As explained in the grooming section above, services with a **large user base** can pose a risk of grooming. They can therefore also be considered as a risk 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; in some cases, such content may be considered CSAM. **Gender, disability** and **sexual orientation** may also be risk factors in CSEA offences leading to the production of CSAM. Due to their significance as a risk factor for CSAM, **large services** and **user base demographics** are included as a general risk factor in the risk profiles.

Functionalities and recommender systems risk factors:

Several functionalities can enable the perpetration of CSAM offences. The functionality of **group messaging** is a risk factor for CSAM, as it allows CSAM to be shared or traded within communities of users. **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 abusers to share CSAM with less risk of discovery. 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. Due to their role in propagating CSAM offences, the functionalities of **group messaging**, **direct messaging**, **posting content**, **encrypted messaging**, **hyperlinks** and **anonymous user profiles** are included in the risk profiles.

Livestreaming is a risk factor as it allows abusers to create CSAM during livestream sessions or from SGII, which can then be widely distributed. This functionality is also included in the risk profiles. Livestreaming is particularly risky when combined with storage and screen capture functionalities, such as the ability to take **screenshots or recordings.**

The ability to **post goods or services for sale** can be used to facilitate the distribution of CSAM. **Cryptocurrency payments** pose a growing threat, as using cryptocurrencies enables offenders to buy CSAM anonymously and evade detection.

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

Functionalities allowing users to **download** CSAM enable users to save local copies of content on their devices, thereby committing the offence of possessing CSAM.

Content **recommender systems** are also a risk factor in the viewing and discovery of CSAM. It is possible that 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/or remove CSAM from their services.

There is also evidence that services which focus on and **emphasise growth** may deprioritise safety measures. 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

6C.107 This section is an overview which looks at how child sexual abuse material (CSAM) manifests online, and how users may be at risks of harm.

6C.108 The scale of CSAM online is rising exponentially, although it is difficult to quantify the exact volume. In 2022, the Internet Watch Foundation (IWF) confirmed that 255,588 reports it received contained CSAM, links to CSAM, or advertised CSAM.[268] And while CSAM can be found on the dark web, 97% of CSAM detected by the Canadian Centre for Child Protection (C3P) Project Arachnid was hosted on the clear web.[269]

6C.109 The severity of the harm is also increasing; more images are being detected in the more severe categories of CSAM. This is particularly true for images of babies and toddlers. The IWF reported in 2022 that Category A material (images involving penetrative sexual activity, sexual activity with an animal, or sadism) were increasing each year. By the end of 2022, the IWF had 8,730 unique hashes[270] of category A material depicting 0-2-year-olds in its hash database, and 411,458 hashes of category A material in total.[271]

6C.110 Self-generated indecent imagery (SGII) is CSAM made by a child. In recent years, there has been a dramatic increase of SGII and it now accounts for the majority of reports (78%) actioned by the IWF.[272] [273]

[268] Internet Watch Foundation, 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].

[269] Canadian Centre for Child Protection (C3P), 2021. Project Arachnid: Online availability of child sexual abuse material. [accessed 21 August 2023].

[270] IWF state that '*A hash is a type of digital fingerprint, or string value that identifies a picture of confirmed child sexual abuse*.' Source: Internet Watch Foundation, 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].

[271] Internet Watch Foundation, 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].

[272] The IWF found that of the 255,571 webpages it acted on during 2022, 78% were assessed as containing SGII. Many of the SGII reports (64%) received by the IWF in 2022 related to girls aged 11 to 13. Source: Internet Watch Foundation, 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].

[273] Ofcom research found that 4% of 13-17-year-old internet users had seen or experienced '*pressure to send photos or personal information to someone*' online in the past four weeks. Ofcom research also found that 3% of 13-17-year-old internet users had seen or experienced '*sharing of, or threats to share, intimate images without consen*t' in the past four weeks. It is unclear from both research studies whether the pressure on children to send images came from other children or from people whom the children did not know. Ofcom, 2023. Experiences of Using Online Services. [accessed 22 September 2023].

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

6C.111 Online CSAM is material which depicts penetrative sexual activity, non-penetrative sexual activity, or other indecent or prohibited imagery of children.[274] It can include photographic images and videos, as well as other non-photographic material, such as drawings, animations and paedophile manuals. CSAM can also include deepfake imagery and imagery created using generative AI tools accessed through extended reality technologies. CSAM can also be non-image based, and can include material which contains advice about grooming or abusing a child sexually, or that which is an obscene article encouraging the commission of other child sexual exploitation and abuse offences (see paragraphs 6C.126, 6C.158 and 6C.159 for further information).[275]

6C.112 Most of the CSAM detected on U2U services at present is content that has previously been shared, and sometimes reshared, hundreds of thousands of times over a period of many years. 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.[276] Perpetrators will often exchange first-generation CSAM in return for other new material.

6C.113 SGII is 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 the 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.[277] Research has shown how this presents on social media services, and how it is used to initiate and maintain romantic or peer relationships.[278] [279]
- **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.[280] In these cases, the negative impacts on the child in the imagery are significant and extend to mental health

[274] Crown Prosecution Service, 2020. Indecent and Prohibited Images of Children. [accessed 18 August 2023].

[275] For more information on what constitutes CSAM see the Illegal Content Judgement Guidance (ICJG - Volume 5, Chapter 26).

[276] 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).

[277] It is crucial for services to note that consensual SGII is driven by children, to 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.

[278] 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.; Thorn, 2022. Online Grooming: Examining risky encounters amid everyday digital socialization. Findings from 2021 qualitative and quantitative research among 9-17-year-olds.

[279] 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. Source: internet matters.org (Katz, A. and El Asam, A.), 2020. Look at Me: Teens, sexting and risks. [accessed 18 August 2023].

[280] 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].

problems[281] and negative social repercussions,[282] especially for girls, who will often experience bullying, harassment, social exclusion and victim-blaming.[283]

- **Aggravated SGII** is imagery of child sexual abuse obtained from a child by a perpetrator, usually as part of the grooming process (as explored in the Grooming section above), 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 sexually abusing themselves and/or others when creating and sending SGII. Many victims experience feelings of self-blame, negative psychological health, and the anxiety of knowing that the images are online.[284]

6C.114 There is growing evidence that extended reality technologies are being exploited for the commission of CSAM offences. These include virtual reality (VR), augmented reality (AR), generative AI (including deepfakes) and CGI experiences. Although, at present, a limited number of services have such functionalities or technologies enabling the creation of CSAM, evidence suggests that this is a fast-growing area. In addition, perpetrators can use user-to-user services to upload, share and sell CSAM created using these technologies.

6C.115 CSAM can be distributed on any service using functionalities that allows users to share images, videos, and text. Visual CSAM (images and videos) as well as CSAM URLs can be disseminated using a number of functionalities such as posts, comments and messages, as well as content tags and descriptions. Moreover, CSAM's high online prevalence and demand means that it can be found on seemingly innocuous services.[285] Perpetrators are likely to choose services for their ordinary appearance, and are attracted to services that they perceive as having weaker content moderation, or lacking robust trust and safety procedures or other features, such as user reporting.

6C.116 In addition to sharing CSAM directly on services, perpetrators also share links among themselves. Large virtual communities of offenders have been seen to share millions of items of CSAM indirectly via hyperlinks.[286] The links are often shared on forums and social media services, directing users to a file-storage and file-sharing service that allows users to upload and share imagery.

6C.117 Perpetrators actively identify and share legal content that is linked to CSAM through various tactics to indicate that they are able to share CSAM, but without uploading it for public view. Content of this type is known as 'contextual CSEA'. It includes the sharing of personal information of CSAM victims and survivors in order to locate material, and sharing contextual images taken from a sexual abuse 'series' to indicate possession of illegal

[281] 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.

[282] 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.

[283] 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 Studie*s, 31(2), pp.243 – 261.

[284] Refer to the Risks of harm to individuals sections for both CSAM and grooming.

[285] See Risks of harm to individuals presented by child sexual abuse material offences for more information. Source: Lee, H. E., Ermakova, T., Ververis, V., & Fabian, B., 2020. Detecting child sexual abuse material: A comprehensive survey. *Forensic Science International: Digital Investigation*, 34, 301022; NCMEC, 2022. 2022 CyberTipline Reports by Electronic Service Providers. [accessed 23 August 2023].

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

material.[287] As with similar perpetrator behaviour for other online harms, this is referred to as 'breadcrumbing'; perpetrators signposting other perpetrators to CSAM.

6C.118 Livestreamed child sexual exploitation and abuse is a known problem, both in the UK and globally.[288] By using screen capturing and recording functionalities, the livestreaming of CSEA can be used to create CSAM. Cases of livestreamed child sexual abuse have often been identified as being streamed from South-East Asia,[289] with perpetrators in Western countries (including the UK),[290] accessing the material, generally in exchange for payment.

6C.119 Being sexually abused has a profound and long-lasting impact on a child. The existence, sharing and viewing of images and videos of that abuse can become a continuing source of trauma for victims and survivors of online CSAM. Many describe feeling constantly in fear, and vulnerable, because their abuse exists as a permanent record online which others can view.[291]

6C.120 Survivors can be re-victimised through reliving the experience of their sexual abuse if they see the material online. In a survey conducted by Canadian Centre for Child Protection (C3P), 69% of victims and survivors indicated that they worried constantly about being recognised, and almost a third (30%) had been identified online or in person by someone who had seen images of their abuse.[292] 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.[293]

6C.121 CSAM also has a broader impact on the population. Six per cent of British adults report having been exposed to CSAM online.[294] The unintentional viewing of CSAM by adults is also likely to cause considerable distress. It may also cause individuals to become desensitised to the material and fall into more regular, intentional viewing that they then find difficult to

[287] Canadian Centre for Child Protection, 2019. How we are failing children: Changing the Paradigm. [accessed 23 August 2023].

[288] The WeProtect Global Alliance outlines that the livestreaming of child sexual exploitation and abuse 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."* Source: WeProtect Global Alliance. Livestreaming child sexual exploitation and abuse.[accessed 25 August 2023].

[289] 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. Source: Internet Watch Foundation, 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].*

[290] 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].

[291] 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.

[292] Sample consisted of a 150 victims and survivors. Canadian Centre for Child Protection, 2017. Survivors' Survey: executive summary 2017. [accessed 25 August 2023].

[293] 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].

[294] The figure is higher for young adults, with 14% of 18-24-year-olds reporting having been exposed to CSAM online. Source: Internet Watch Foundation, 2022. More than one in 10 British young people exposed to online child sexual abuse. [accessed 25 August 2023].

stop.[295] A survey conducted by the Finnish organisation, Protect Children, with people who view CSAM, found that 50% of respondents[296] wanted to stop viewing CSAM.[297]

6C.122 Tackling CSAM online can disrupt future sexual abuse through physical contact. Research suggests that perpetrators of contact child sexual abuse are likely to have accessed CSAM online before offending in person.[298] Protect Children found that thirty-seven per cent of surveyed individuals who had viewed CSAM tried to seek direct contact with a child afterwards.[299]

6C.123 Children may also unintentionally view this material, and this can have a significant impact on their wellbeing. A Childline report using data from UK counselling sessions found that young people who had unintentionally accessed CSAM online were reluctant to tell anyone about it, fearing that they would either not be believed, or that they might even be arrested. 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*".[300]

Evidence of risk factors on user-to-user services

6C.124 We consider that the risk factors below are liable to increase the risks of harm relating to CSAM. This is also summarised in the grey box at the beginning of the chapter.

6C.125 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.

6C.126 The evidence provided in the CSAM assessment below focuses on image-based content. References to CSAM could include content which constitutes a 'paedophile manual' (i.e. an item that contains advice or guidance about abusing children sexually) or that which is an obscene article tending to deprave and corrupt others by encouraging them to commit a specified offence relating to CSAM or another form of CSEA. These items or articles may contain images or videos, but need not be image-based. However, the findings about the creation, uploading, sharing and distribution of image-based CSAM are often likely to be relevant to these kinds of content as well.

[295] 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*].

[296] Sample of 3,935.

[297] 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].

[298] Independent Inquiry into Child Sexual Abuse (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].

[299] 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].

[300] NSPCC, 2016. Online child sexual abuse images: Doing more to tackle demand and supply. [accessed 25 August 2023].

Risk factors: Service types

6C.127 While any service that allows users to share images or 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: social media services, video-sharing services, discussion forum and chat room services, file-storage and file-sharing services, private messaging services, and online adult services.

Social media services

6C.128 There is evidence that social media services are known high-risk spaces for CSAM sharing. This may be in part because some social media services have a large user base. For instance, the NCMEC 2022 CyberTipline report shows that Facebook, Instagram, and Omegle (alongside Google and Whatsapp), accounted for 90% of all reports.[301] Other well-known social media services also show reports into the tens or hundreds of thousands.

6C.129 Social media services can be used as directories that allow perpetrators to signpost users to other services where CSAM is shared.

6C.130 Certain functionalities on social media services have been found to be risk factors. There is evidence of hyperlinks and plain-text URLs to CSAM being posted on social media services in a 'scattergun' approach that can drive up web traffic, and income for those hosting CSAM content.[302]

Video-sharing services

6C.131 User comments on some video-sharing services can be used to 'breadcrumb' and thereby facilitate access to CSAM.[303] Livestreaming, a functionality that is common to many video-sharing services, has also been found to be a risk factor, particularly when paired with screen capture tools.[304]

Discussion forums and chat-room services

6C.132 Discussion forums can be used to embed and advertise CSAM material. The IWF reported that 5% of the child sexual abuse imagery it detected was on forums.[305] In a separate study in 2018, specifically examining captures of livestreamed child sexual abuse, the IWF found that 73% 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*".[306] [307]

[301] NCMEC, 2023. CyberTipline 2022 Report. [accessed 22 September 2023].

[302] Internet Watch Foundation, 2022. Public warned as 'disturbing' new trend risks exposure to child sexual abuse material online. [accessed 25 August 2023]. See Risk factors: functionalities and recommender systems for more information.

[303] WIRED (Orphanides, K. G.), 2019. On YouTube, a network of paedophiles is hiding in plain sight. [accessed 25 august 2023]. See Risk factors: functionalities and recommender systems for more information..

[304] See Risk factors: functionalities and recommender systems for more information..

[305] Internet Watch Foundation, 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].

[306] Based on a sample of 2,082 images over three months. Source: Internet Watch Foundation, 2018. Trends in Online Child Sexual Exploitation: Examining the distribution of livestreamed child sexual abuse. [accessed 22 September 2023].

[307] 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.

File-storage and file-sharing services

6C.133 File-storage and file-sharing services, in particular services that allow users to upload and share access to images, present a significant risk for hosting CSAM. Perpetrators can store CSAM on these services and distribute URLs directing users to these collections.[308]

6C.134 According to Dorotic and Johnsen, perpetrators *"use file hosting platforms to store and distribute CSAM"*. This is done through URLs which are shared to services such as image boards, offender forums and 'chats'.[309] In 2021, INHOPE, a global network of organisations working to tackle CSAM, reported that approximately 26% of the CSAM it detected was hosted by file hosts and 25% was hosted by image hosts.[310]

6C.135 The IWF found that 85% of images or videos detected of livestreamed child sexual abuse were stored on an 'image-hosting service'.[311] 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".[312] [313]

6C.136 Other examples of file-storage and file-sharing services such as 'cyberlockers' and 'image stores' made up 3% and 10% of cases respectively of the child sexual abuse imagery that the IWF reviewed in 2022.[314]

Messaging services

6C.137 Messaging services have also been found to be a risk factor. Messaging services have been identified as one of the online channels that perpetrators use to distribute CSAM.[315] Research by Interpol[316] 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.[317]

6C.138 Messaging can be used in conjunction with other services to facilitate CSAM offences; for example, to communicate with victims or facilitators of livestreamed CSEA.[318] Some offenders start this interaction via messaging, then move to another service to watch the livestream.[319]

[308] We understand that perpetrators may be more likely to choose these services if they are encrypted, or if access is time-limited.

[309] 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].

[310] InHope, 2022. Annual Report 2021. [accessed 25 August 2023].

[311] Further, in 2017, 69% of the child sexual abuse images actioned by the IWF was hosted on image host websites. Internet Watch Foundation, 2018. Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse. [accessed 22 September 2023].

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

[313] The Internet Watch Foundation 2022 report found that 77% of child sexual abuse imagery appeared on imaging hosts. Source: Internet Watch Foundation, 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].

[314] Internet Watch Foundation, 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].

[315] 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 child sexual abuse material offences for more information.

[316] Interpol is an international organisation that facilitates police cooperation on international crime.

[317] Interpol, 2020.

[318] 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].

[319] 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].

6C.139 Messaging services also offer direct messaging, which can be used by offenders to engage with one another and share CSAM.[320] This is also true of group messaging. More detail can be found below in the Risk factors: Functionalities and recommender systems section. Private messaging services with encryption are particularly risky, as they make the exchange of CSAM harder to detect.[321]

Adult services

6C.140 Evidence indicates that perpetrators are using adult services to disseminate CSAM. A 2021 report by C3P's Project Arachnid stated that the ESP Serverel, which is the hosting provider for at least 1,200 unique websites sharing ephemeral 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).[322]

6C.141 Perpetrators are also using online adult services to seek out CSAM. This is evidenced by the chatbot released on Pornhub, which is used to intercept searches of 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.[323]

Risk factors: User base

User base size

6C.142 The size of a service's user base is a possible risk factor; 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'.[324]

6C.143 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 in order to exploit such functionalities.

User base demographics

6C.144 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.

[320] See Risk factors: functionalities and recommender systems for more information.
[321] See Risk factors: functionalities and recommender systems for more information.
[322] Canadian Centre for Child Protection, 2021. Project Arachnid: Online availability of child sexual abuse material. [accessed 22 September 2023].
[323] Internet Watch Foundation, 2022. Internet Watch Foundation, Stop It Now, and Pornhub launch first of its kind chatbot to prevent child sexual abuse. [accessed 30 August 2023].
[324] NCA, 2021. National Strategic Assessment of Serious and Organised Crime. [accessed 31 August 2023].

6C.145 Data suggests that user base characteristics including **age**, **gender, ethnicity, disability, sexual orientation and media literacy** could lead to an increased risk of harm to individuals.

6C.146 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 self-generated indecent imagery.

6C.147 Gender is a risk factor for CSAM. The demographics of children depicted in CSAM are skewed towards girls; 96% of the reports processed by the IWF in 2022 depicted exclusively girls, but content depicting boys tended to be of higher severity (based on CSAM category) than content depicting girls.[325] In terms of offender demographics, the IWF found that of the CSAM they analysed, where a perpetrator was visible, they tended to be male.[326] Other studies also point to most CSAM perpetrators being male; one study by the United States Sentencing Commission found that 94.3% of offenders involved in CSAM production were male.[327]

6C.148 Evidence on ethnicity as a risk factor is inconclusive, as evidence finds 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.[328] In terms of other ethnicities, 10.1% of children were classified as Hispanic or Latinx, 9.9% were Asian, and 2.1% were Black. For perpetrators, the research found that 12.2% were Hispanic or Latinx, 4.2% were Black and 3.2% were Asian.

6C.149 The gender and socio-economic background of users has been found to increase the risk of SGII occurring. Evidence has found that girls in general, as well as girls from less privileged backgrounds, are at greater risk of experiencing the non-consensual sharing of SGII.[329]

6C.150 There is evidence that disability can be a risk factor in the creation of CSAM. Research indicates that perpetrators target and exploit the vulnerability of children with disabilities in order to sexually abuse them,[330] 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.[331] In particular, research has found that children with disabilities or who are neurodivergent are more vulnerable to pressures from others to produce SGII.

[325] Internet Watch Foundation, 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].

[326] Internet Watch Foundation, 2022. The Annual Report 2021. [accessed 30 August 2023].

[327] United States Sentencing Commission, 2021. Federal Sentencing of Child Pornography: Production Offenses. [accessed 22 September 2023].

[328] ECPAT and Interpol, 2018. Towards a global indicator on unidentified victims in child sexual exploitation material: summary report. [accessed 30 August 2023].

[329] Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 30 August 2023].

[330] Independent Inquiry Child Sexual Abuse, 2022. Child sexual exploitation by organised networks: Investigation Report. [accessed 30 August 2023].

[331] Vera Institute of Justice (Smith, N. and Harrell, S.) 2013. Sexual Abuse of Children with Disabilities: a National Snapshot. [accessed 30 August 2023].

6C.151 As with disability, there is evidence that children who are not heterosexual, or who are questioning their sexual identity, may be more vulnerable to pressure from others to produce SGII.

6C.152 The more sophisticated perpetrators of CSAM offences are likely to have higher media literacy, which may enable them to evade detection and find new ways of producing, accessing and sharing CSAM online.

Risk factors: Functionalities and recommender systems

User identification

User profiles and fake user profiles

6C.153 User profiles can be used in different ways to facilitate CSAM offences, including the use of user profiles that impersonate victims and survivors in order to signpost to CSA material.

6C.154 Fake user profiles can be created by perpetrators to impersonate victims and survivors. The NSPCC has warned of 'tribute sites' and 'tribute user profiles', that impersonate victims and survivors. They are used by abusers so that they can connect with each other and share advice.[332] This is a CSEA breadcrumbing technique that facilitates the commission of CSAM offences.

6C.155 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.[333] The use of these kinds of services by perpetrators is further evidenced by the existence of a network of young users who collect and share the usernames of users who have committed sexual misconduct or engaged in 'creepy interactions'.[334]

6C.156 Moreover, 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.

Anonymous user profiles

6C.157 Services that allow users to create anonymous user profiles and allow 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 was identified in a study into technologies used to commit child sexual abuse offences: 82% of the offenders surveyed indicated that anonymity was of at least moderate importance for conducting CSAM offences.[335] From this study we can assume that services which offer perceived anonymity are attractive to offenders. They may be especially 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.[336]

[332] NSPCC response to Ofcom 2022 Call for evidence: First phase of online regulation.

[333] Cox, J, 2018. December 6. TikTok, the app super popular with kids, has a nudes problem. *VICE,* 6 December. [accessed 20 August 2023].

[334] Broderick, R., 2019. TikTok Has A Predator Problem. A Network Of Young Women Is Fighting Back, *Buzzfeed,* 26 June. [accessed 30 August 2023].

[335] 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).

[336] In such cases, the identity of users may also be unknown to services.

User networking

User groups

6C.158 User groups are places where like-minded individuals can share content and engage with one another.[337] Perpetrators take advantage of these spaces and connect in online communities to exchange CSAM, as well as to give advice regarding abusive behaviours and practices. The NSPCC notes 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.[338]

6C.159 In some user groups, perpetrators share CSAM with one another and thereby commit a CSAM offence, as well as facilitate CSAM offences through the exchanging of ideas, advice and tradecraft tips. Sharing CSAM is sometimes a condition of entry into some closed user groups.

User communications

Livestreaming

6C.160 Livestreamed child sexual exploitation and abuse 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.[339] By using screen capturing and recording functionalities, the livestreaming of CSEA can be used to create CSAM.

6C.161 Livestreaming can happen anywhere in the world. A common methodology involves a perpetrator in a developed nation, who takes advantage of economic disparity by paying for the live sexual abuse of children in less-developed nations. Interpol predicted that this type of child sexual exploitation would increase due to the Covid-19 pandemic, because travel restrictions limited abusers' access to young people, and Covid-19 worsened the economic conditions for many people worldwide.[340] Subsequently, there was a 265% increase in recorded cases of child sexual exploitation and abuse livestreamed from the Philippines during the quarantine period March to May 2020**.**[341]

6C.162 The Internet Watch Foundation's report on livestreaming captures analysed over 2,000 indecent images of children taken from livestreams. Ninety-eight per cent of these showed children who appeared to be 13 years or under.[342]

6C.163 In recent years there has been a significant increase in livestreamed child sexual abuse, including self-generated content. Globally, the UK is estimated to be the third largest consumer of livestreamed child sexual abuse.[343]

6C.164 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

[337] Chiang, E., Nguyen, D., Towler, A., Haas, M. and Grieve, J., 2020. Linguistic analysis of suspected child sexual offenders' interactions in a dark web image exchange chatroom, *The International Journal of Speech, Language and the Law*, 27 (2).
[338] NSPCC, 2019. Online abuse: learning from case reviews. [accessed 30 August 2023].
[339] Offences relating to child sexual exploitation are discussed in chapter 6J: Unlawful immigration and human trafficking
[340] Interpol, 2020. Threats and trends Child sexual exploitation and abuse: covid-19 impact. [accessed 22 September 2023].
[341] WeProtect Global Alliance, n.d. Live Streaming Child Exploitation and Abuse. [accessed 22 September 2023].
[342] Internet Watch Foundation, 2018. Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse. [accessed 22 September 2023].
[343] Independent Inquiry into Child Sexual Abuse, 2020. The Internet: Investigation Report. [accessed 22 September 2023].

found evidence that children who livestream are sometimes asked to perform sexual acts.[344] Of those children who livestream,[345] 6% had received requests to change or remove their clothes.[346]

Direct messaging, encrypted messaging, and ephemeral messaging

6C.165 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 private messaging services or 'message applications' during the Covid-19 pandemic in 2020, which were centred around direct messaging.[347]

6C.166 Encrypted messaging makes the exchange of CSAM hard to detect.[348] 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 makes 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.[349]

6C.167 Because of the difficulty in detecting CSAM over encrypted messaging, perpetrators are likely to seek out spaces with encrypted messaging in order to disguise their activity.

6C.168 Ephemeral messaging functionalities also make it harder for CSAM to be detected. Ephemeral messaging can also 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.

6C.169 Direct messaging, particularly 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.[350]

Group messaging

6C.170 Group messaging functionalities can also be used to share CSAM. A study by Steel *et al*. in 2020 noted cases where abusers traded and shared CSAM in group chats on messaging applications.[351] It is therefore possible that services with similar private group chat facilities could facilitate CSAM offences.

[344]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.

[345] 24% of all children have done a livestream broadcast. Source: NSPCC, 2018. Livestreaming and video-chatting. [accessed 22 September 2023].

[346] NSPCC, 2018. Livestreaming and video-chatting. [accessed 22 September 2023].

[347] Interpol, 2020. Threats and trends child sexual exploitation and abuse: COVID-19 impact. [accessed 22 September 2023].

[348] 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.

[349] Internet Watch Foundation, 2022. Not all Encryption is the same: social media is not ready for End-to-End Encryption. [accessed 30 August 2023].

[350] Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 22 September 2023].

[351] 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.

6C.171 Group messaging encompasses 'chat functions' which can be used to share CSAM URLs. Perpetrators tend to operate within groups or networks which trade content with each other; as such, allowing users to communicate through chat functions can provide a space for these networks to operate and share CSAM URLs with one another. Services may also 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.

6C.172 Group messaging functionalities can be used to non-consensually share SGII to large groups of people. 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.[352] 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.[353]

Commenting on content

6C.173 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 facilitate perpetrators' access to CSAM.[354]

Posting content (text, images, video)

6C.174 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 (images and videos), and links or URLs to CSAM, on both open and closed channels of communication.

6C.175 For example, 77% of the CSAM reports dealt with by the Internet Watch Foundation in 2022 were from services which hosted images.[355] 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

6C.176 The use of CSAM within an advert is a CSAM offence in itself, as well as enabling the commission of further CSAM offences.

6C.177 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

6C.178 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,

[352] Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 22 September 2023].

[353] Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 22 September 2023].

[354] WIRED (Orphanides, K. G.) 2019. On YouTube, a network of paedophiles is hiding in plain sight. [accessed 22 September 2023].

[355] "*These sites provide 'storage' for images which either appear on dedicated websites or are shared within forums*". Source: Internet Watch Foundation, 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].

can enable offenders to buy CSAM anonymously and evade detection. The IWF notes that the number of websites found to accept cryptocurrency payments for CSAM has doubled in most years since 2015.[356] Other potential payment options for CSAM include direct payment mechanisms, such as credit card or money transfer services.

Content exploring

Hyperlinking

6C.179 Hyperlinks facilitate access to CSAM as they can be used to direct abusers to CSAM hosted on third-party sites. NSPCC refers to this as 'digital breadcrumbing', where abusers use services to signpost other abusers to CSAM hosted on other sites.[357] This includes the use of QR codes,[358] working in a similar way as hyperlinks.

6C.180 URLs – both in the form of hyperlinks and plain text – can be shared between individual offenders 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 of otherwise legitimate services.

6C.181 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.[359]

Building lists or directories

6C.182 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

6C.183 The ability to search for user-generated content within a U2U service is a risk factor. Autocomplete suggestions in a U2U service's search box could 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.'[360]

6C.184 In addition, some offenders use search functions on adult services to search for terms associated with CSAM.

Content editing

Editing visual media

6C.185 The ability to edit images and videos can be used to artificially change legal pornography to make participants appear as children. Editing functionalities can also be used by abusers to edit CSAM to evade detection technologies such as cryptographic hash matching.

[356] Internet Watch Foundation, 2022. Websites offering cryptocurrency payment for child sexual abuse images 'doubling every year'. [accessed 22 September 2023].

[357] 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].

[358] NSPCC response to Ofcom 2022 Call for evidence: First phase of online safety regulation.

[359] Internet Watch Foundation, 2022. Public warned as 'disturbing' new trend risks exposure to child sexual abuse material online. [accessed 30 August 2023].

[360] Hern, A., 2018. Facebook apologises for search suggestions of child abuse videos, *The Guardian*, 16 March. [accessed 30 August 2023].

6C.186 There are examples of generative AI being used by perpetrators to edit images and videos to produce CSAM. In particular, the creation of deepfake[361] material (see chapter 6M – intimate image abuse for further information) is a relatively new trend, but there is evidence that this technology is being used by perpetrators to create deepfake CSAM.[362] The use of deepfakes in CSAM production is still very new, and the resulting images and videos are hard to identify.

Content storage and capture

Downloading content

6C.187 The ability to download content, including CSAM, allows users to save local copies to their computers and devices, thereby committing the offence of possessing CSAM. This functionality enables perpetrators to build very large collections of CSAM. The National Crime Agency has found some perpetrators who have downloaded over a million child sexual abuse images to their devices.[363]

Screen capturing or recording

6C.188 There is evidence of perpetrators discussing the technique of 'capping' – the capturing of live images and videos.[364] This involves both capturing images from livestreams and capturing images of SGII videos from victims who have been groomed and coerced.

6C.189 A study by the IWF found that screen captures of livestreaming can be used by offenders to produce CSAM, which can then be distributed to other sites online.[365] 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; 88% 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.

Recommender systems

Content recommender systems

6C.190 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.

[361] 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.

[362] Hedgecoe, G., 2023. AI-generated naked child images shock Spanish town of Almendralejo, *BBC News*, 24 September. [accessed 27 September 2023].

[363] For example: Luck, F., 2023. Former GP caught with 1.2m indecent images of children jailed. *BBC News*, 23 February. [accessed 30 August 2023].

[364] 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.

[365] Internet Watch Foundation, 2018. Trends in Online Child Sexual Exploitation: Examining the distribution of lives-streamed child sexual abuse. [accessed 22 September 2023]

6C.191 There is evidence of users being recommended inappropriate, but not necessarily illegal, content.[366] 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.[367]

Risk factors: Business models and commercial profiles

Revenue models

Advertising-based model

6C.192 Advertising features on user-to-user services can be used by perpetrators of CSEA. Research by IWF, commissioned by the Home Office, found that legitimate adverts can be inadvertently used to fund websites 'dedicated to child sexual abuse'[368] 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. 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.[369]

6C.193 A report by WIRED found that a prominent online service had put adverts from major brands alongside video content of young children which attracted offenders.[370] The report states that "*videos that are seemingly popular with other paedophiles, most of which have hundreds of thousands of views and dozens of disturbing comments*"[371] had adverts for major brands.

Commercial profile

Low-capacity and early-stage services

6C.194 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 (e.g. 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 in order to share and view CSAM undetected.

Growth strategy

6C.195 Evidence suggests that services which prioritise and emphasise growth (e.g. prioritising user growth) may put insufficient resources towards effectively moderating harmful content and preventing offenders from exploiting the service.

[366] 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].
[367] 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.
[368] Home Office, 2018. Advertisers urged to help tackle online child sexual exploitation. [accessed 22 September 2023].
[369] Home Office, 2018. Advertisers urged to help tackle online child sexual exploitation. [accessed 22 September 2023].
[370] WIRED (Orphanides, K. G.) 2019. On YouTube, a network of paedophiles is hiding in plain sight. [accessed 22 September 2023].
[371] WIRED (Orphanides, K. G.) 2019. On YouTube, a network of paedophiles is hiding in plain sight. [accessed 22 September 2023].

6D. Encouraging or assisting suicide (or attempted suicide) or serious self-harm offences

Warning: This chapter contains discussion of suicide, self-harm and eating disorders.[372]

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 self-harm or end their life. 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.

Content related to suicide and self-harm is extremely sensitive; while there may be users who share this content to cause harm to others, some users may share this content to find supportive communities or to reflect their own experiences as part of a healing process. It can include users who suffer with suicidal or self-harm ideation, 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 qualitative information for insights into risk factors.

Service type risk factors:

Discussion forum and chat room services can act as spaces where suicide and 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 suicide and self-harm. This is particularly true of services that facilitate discussions among smaller groups of users. 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 or self-harm 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. Due to their role in propagating this offence, discussion forum and chatroom services and social media services are included in the risk profiles.

Video-sharing services can also enable views to watch suicide and self-harm content, particularly through livestreaming.

[372] 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]().

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 suicide, self-harm and eating disorders.

User demographics can play a significant role in the risk of physical or psychological harm that can occur from content that encourages or assists suicide. Specifically, users who are 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 suicide-related purposes, and to be susceptible to copycat behaviour.

Functionalities and recommender systems risk factors:

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. This functionality is included in the risk profiles.

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 suicide and self-harm content. Due to their role in sharing suicide and self-harm content, the functionalities of posting content and re-posting or forwarding content have been included in the risk profiles.

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 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. Due to the risk of distributing suicide or self-harm content, content recommender systems are included in the risk profiles.

Some functionalities intersect to create high-risk context. **Livestreaming** is a risk factor that has been used to share real-time acts of suicide or self-harm. Livestreams are often paired with **user groups,** allowing users to communicate with one another and/or leave **comments on content**, which can be used to encourage the suicide or self-harm depicted on the livestream. More generally, suicide- or self-harm-related messaging can be found in the comments sections on posted content. These three functionalities (livestreaming, user groups, commenting on content) are included in the risk profiles for their role in enabling suicide and self-harm content.

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

6D.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the suicide and serious self-harm 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’).

6D.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.[373]

6D.3 Suicide and self-harm are not in themselves criminal offences, and an individual attempting to do this will not be penalised. Individuals suffering with their mental health, including thoughts of suicide and self-harm, should seek support without fear of negative consequences.

6D.4 Services should be aware that suicide or self-harm content varies and is not always shared with malicious intent. Users who share suicide or self-harm content may be suffering from

[373] As with other chapters, we have considered evidence of suicide and/or 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 these offences.

mental health problems themselves, and use online spaces to express their feelings and seek support by connecting with others who may be having similar experiences. 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.

6D.5 We have considered self-harm and suicide-related evidence side by side 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.[374] For the purposes of our assessment, including to assess the impact of characteristics on the risks of harm, we also treat some suicide and 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' below).

6D.6 Where data from Ofcom's Online Experience Tracker has been included below, 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

6D.7 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. In regard to suicide and self-harm, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

6D.8 In this chapter, we consider the priority offence of encouraging or assisting suicide.[375] In addition, we consider the relevant non-priority offence of encouraging or assisting serious self-harm.[376]

6D.9 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6D.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.[377]

6D.11 The new 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 (GBH).[378]

6D.12 For both offences, 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 suicide, attempted suicide or actual self-harm to amount to illegal content.

[374] 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].

[375] 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.)).

[376] Section 184 of the Online Safety Act 2023.

[377] Samaritans, 2020. Understanding self-harm and suicide content online. [accessed 24 May 2023]

[378] 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.

6D.13 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** (Volume 5, Chapter 26).[379]

How encouraging or assisting suicide or serious self-harm offences manifest online

6D.14 This section is an overview which looks at how the offences of encouraging or assisting suicide or serious self-harm manifest online, and how users may be at risk of harm.

6D.15 To put the risks of harm from this offence into context, Ofcom's Online Experiences Tracker (OET) found that 4% of UK internet users had seen or experienced content 'related to self-harm or suicide' in the past four weeks.[380] Younger adults are more likely to claim to have seen content related to self-harm or suicide, with those aged 18-34 more likely than the average internet user to claim experience of this content (7% vs 4%).[381]

6D.16 While it is challenging to quantify the social and economic cost of this harm, the Department for Health estimated that the cost of each suicide was £1.67m in 2009 prices[382] (£2.23m in 2022/23 prices).[383]

Risks of harm to individuals presented by the offences of encouraging or assisting suicide or serious self-harm online

6D.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.[384] [385] 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.[386]

6D.18 There are at least two distinct groups of users who are likely to be at risk: those who encounter this content by accident (e.g. when searching for content that overlaps with hashtags to share harmful self-harm content), 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.

[379] 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.

[380] This is likely to include content that could be deemed illegal.

[381] Ofcom, 2023. Online Experiences Tracker 2021-2022.Waves 1 and 2 (18+)– base size n=4334 (those who have encountered self-harm or suicide content online in the past 4 weeks).

[382] Department of Health, 2016. Department of Health written evidence to Parliament Health Select Committee 2016. [accessed 27 September 2023].

[383] Ofcom price update using GDP deflator.

[384] Samaritans, 2022. Towards a suicide-safer internet. [accessed 24 May 2023]

[385] These examples may not necessarily be illegal in all cases. For information on what could be considered illegal, please refer to the ICJG guidance.

[386] 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.

6D.19 Studies find it challenging to establish a causal relationship between exposure to online suicide or self-harm content alone, and its effect in increasing the risk of self-harm or suicidal thoughts.[387] However, two-thirds (66%) of UK adults say they are concerned about the accessibility of harmful suicide or self-harm content online.[388] This is likely to include content that could be considered illegal.

6D.20 Several studies have further explored the impact of exposure to self-harm and suicide content. A US-based study of 18-29-year-olds found that those who view content encouraging self-harm on a social media service, either intentionally or by accident, are at a higher risk of self-harm or suicide.[389] Moreover, 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 presented harmful effects,[390] including being 'triggered' by the images, which may lead to normalising or escalating self-harm through sharing tips and ideas, and being encouraged to share images or compete with others.[391]

6D.21 Suicide or self-harm content may have a 'contagion' effect. The Royal College of Psychiatrists identify *"the well-known 'contagion' effects of self-harm in inpatient units".*[392] This could also apply to online contexts. 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).[393]

6D.22 Users' mental state at the time of using the internet may also influence the impact that certain types of suicide or self-harm content have 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.[394]

[387] 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].

[388] 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].

[389] 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].

[390] 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].

[391] 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].

[392] Royal College of Psychiatrists, 2020. Technology use and the mental health of children and young people. [accessed 10th July 2023].

[393] Samaritans, 2022. Towards a suicide-safer internet. [accessed 24 May 2023]

[394] 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].

6D.23 The potential negative effects of this content were also evident in a national survey by Swansea University and Samaritans (where 87% of the sample[395] 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.[396] 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%.

6D.24 Content that encourages eating disorders, which can carry substantial risks of harm, could potentially be linked to these offences.[397] Eating disorders (EDs) are associated with adverse physical consequences and suicide risk, and anorexia nervosa has the highest mortality rate of any mental illness.[398] 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 encountered content which they described as harmful in the context of their eating disorder.[399]

6D.25 Exposure to 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.[400] Use of pro-ED websites has also been shown to discourage help-seeking and maintain or prolong an eating disorder.[401] As early intervention is key to effective treatment for an eating disorder, delays to help-seeking and treatment are significant.[402] [403]

6D.26 Some evidence suggests that 13-17-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.[404] [405] Indeed, it has been estimated that 25% of people with

[395] 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].

[396] Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. [accessed 10 July 2023].

[397] 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].

[398] 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].

[399] Beat, 2023. Online Safety and Eating Disorders, [accessed 10 July 2023].

[400] 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].

[401] 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]

[402] Beat, 2022. Best practice in ensuring early intervention for eating disorders. [accessed 10 July 2023].

[403] 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].

[404] 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.

[405] 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].

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.[406] [407] [408]

6D.27 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.*[409] Users will also occasionally request 'meanspo' from other users, as a form of inspiration and enforcement for their disordered eating.[410]

Evidence of risk factors on user-to-user services

6D.28 We consider that the risk factors below are liable to increase the risks of harm relating to encouraging or assisting suicide or serious self-harm. This is also summarised in the grey box in the introduction.

Risk factors: Service types

6D.29 Research indicates that the following types of services can be used to commit or facilitate offences related to suicide and 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

6D.30 Our evidence shows that discussion forums and chat room services can act as spaces where suicide and self-harm 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 harmful or distressing to users.[411]

[406] 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].

[407] 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].

[408] 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].

[409] Wales Online, 2022. Coaches are training children to be anorexic with vile comments online. [accessed 10 July 2023].

[410] 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].

[411] 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].

6D.31 A number of deaths in the UK in recent years have reportedly involved chatrooms,[412] 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.[413]

6D.32 A small-scale qualitative study in the UK looking at 18-24-year-olds who had previously had suicidal thoughts found that distressing content on 'online chat groups, blogs, and forums' had 'exacerbated their suicidal feelings',[414] although a number of respondents reported that these spaces can also enable users to find help and support. For some respondents, content on online forums emerged as the main factor in generating negative effects.[415]

6D.33 A US study, looking at young people (aged 14-24) 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*."[416] 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.

6D.34 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." Some individuals have reported being encouraged to commit 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.[417]

6D.35 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 which helps them deal 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 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.[418]

[412] 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].

[413] 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.*

[414] 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, *Archives of suicide research: official journal of the International Academy for Suicide Research*, 22 (4). [accessed 10 July 2023].

[415] 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, *Archives of suicide research: official journal of the International Academy for Suicide Research*, 22 (4). [accessed 10 July 2023].

[416] 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].

[417] 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].

[418] 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].

Social media services

6D.36 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,[419] either intentionally or by accident, are at a higher risk of self-harm or suicide.[420] 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.[421] The research also indicates that functionalities such as 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.[422]

6D.37 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.[423]

Video-sharing services

6D.38 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.[424] These livestreams occasionally also contain potentially illegal content within the comment threads (see Commenting on content for more information).

Services that enable online community building

6D.39 Services that allow users to build online communities are a risk factor, as online communities can act as spaces where eating disorders are promoted or encouraged. A 2021 study was able to demonstrate 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.[425] Similar results have been found in individuals with no history of an ED.[426]

[419] In particular, Instagram, as the focus of this study at the time.

[420] 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'.

[421] 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.

[422] 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].

[423] 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].

[424] See 'livestreaming' in Risk factors: functionalities and recommender systems section for more information.

[425] 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].

[426] 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].

6D.40 While pro-ED communities have been argued to offer a sense of emotional support, this support can depend on unhealthy group conformity.[427] Pro-ED content can also involve images to inspire weight loss (so called 'thinspiration' or 'thinspo').[428] 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.[429]

Risk factors: User base

User base size

6D.41 Services with both large and small user bases pose risks in relation to suicide and self-harm content, for different reasons.

6D.42 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.[430] (See Commenting on content for more information). This in turn heightens the risk of contagion effects, as described earlier. (See Risk factors: functionalities and recommender systems for further information)

6D.43 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.

6D.44 Pro-ED content, for example, can appear on large services as well as on smaller services, websites and blogs.[431] [432] Use of specialised pro-ED websites can expose users to extreme content and present a barrier to eating disorder recovery.[433] [434]

User base demographics

6D.45 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.

6D.46 The data suggests that user base characteristics including **age, mental health, ethnicity, religion**, **sexual orientation** and **gender** could lead to increased risks of harm to individuals.

[427] 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].

[428] 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].

[429] 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] [accessed 11 July 2023].

[430] Ekō, 2023. Suicide, Incels, and Drugs: How TikTok's deadly algorithm harms kids. [accessed 11 July 2023].

[431] 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].

[432] 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].

[433] 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].

[434] 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].

6D.47 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', which may not necessarily include content deemed to meet the illegal threshold.

6D.48 The data suggests that the age of users influences their susceptibility to suicide and self-harm content, as well as a user's likelihood of encountering content related to self-harm or suicide. Ofcom OET research suggests that this content is seen or experienced online significantly more by younger adults, aged 18-24 (10%), than by the average of UK internet user (4%).[435] The evidence also suggests that younger adults are more likely to experience the contagion effect (i.e., copycat behaviour)[436] and to have used the internet for suicide-related purposes (among those who had been in contact with mental health services).[437]

6D.49 Our evidence also show that the mental health of users affects risk levels.

6D.50 OET data suggests that this self-harm or suicide content is experienced significantly more by internet users with mental health conditions[438] (8% vs 3% with no conditions).

6D.51 Other evidence also suggests that those with existing mental health concerns may be more likely to encounter suicide or self-harm content.[439] 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.[440] 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.[441]

6D.52 Additionally, the **ethnicity**, **religion**, **sexual orientation** and **gender** of users may play a role in increasing the risk of harm related to suicide or self-harm content.

6D.53 OET data suggests that this online harm is experienced significantly more by internet users who are of mixed ethnicity (7%), Black (6%) and of multiple ethnicities (6%) compared to those who are white (3%). It also suggests that the online harm is experienced more by Muslim users (6%) than Christian and 'other religion' users (3%) as well as those who identify with 'other' sexuality (11%) and bisexual (10%) vs heterosexual (3%). Finally, females (4%) were more likely than males (3%) to experience this online harm.

[435] Ofcom, 2023. Experiences using online services. [accessed 25 September 2023].

[436] 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].

[437] 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].

[438] Ofcom's OET refers to mental health as 'Anxiety, depression or trauma-related conditions, for example.'

[439] 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].

[440] Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. [accessed 10 July 2023]

[441] Samaritans response to 2023 Ofcom Call for Evidence: Second phase of online safety regulation: Protection of Children

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

6D.54 Fake user profiles can increase the risk of pro-suicide and self-harm content being disseminated on a service. There are case examples where perpetrators have created false identities to maliciously encourage others to take their own lives. False identities can be used to create personas that users would be likely to relate to, and be influenced by.

6D.55 For example, a case in the USA involved a potential perpetrator (described as a 'serial predator') who was convicted for using fake identities to encourage individuals to join bogus suicide pacts and to assist with suicides by suggesting methods for victims and survivors to use.[442]

Anonymous profiles

6D.56 While anonymity has important benefits[443], 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.

6D.57 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/friends if they encountered this suicide-related content.[444]

6D.58 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".*[445]

User networking

User groups

6D.59 User groups that are dedicated to discussing self-harm or 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 self-harm.

[442] 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].

[443] 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.

[444] 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].

[445] 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].

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.[446]

User communication

Livestreaming

6D.60 A livestreaming functionality can increase the risk of users being exposed to self-harm and suicide content. It can also increase the risk of those hosting the livestream, who may be struggling themselves, being exposed to harmful messages.

6D.61 There have been numerous cases of livestreaming functionalities being used to show users self-harming or ending their life in real time.

6D.62 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.[447] Another livestreaming suicide case was that of 43-year-old male, who took his own life while livestreaming on a video chatroom service in July 2018. On this service, users exploited the functionalities to insult, provoke and abuse each other.[448]

6D.63 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.[449]

6D.64 Livestreaming can intersect with group messaging and commenting functionalities to increase the risks of harm. Users can often message one another as a group within the livestream and/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. Group messaging and commenting is explored further in the sections below.

Group messaging

6D.65 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.

[446] 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].

[447] 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].

[448] 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].

[449] 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].

6D.66 In the UK between 2001 and 2008 there were at least 17 deaths involving chatrooms or sites that provide advice on suicide methods.[450] 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).[451]

Commenting on content

6D.67 Potentially illegal suicide or serious self-harm-related messaging can be found in the comments sections on posted content. An Australian study which looked at comment replies to suicide-related posts on 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*".[452]

6D.68 Similarly, a study looked at the comment threads of 26 livestreaming videos where an individual had threatened to take their own life.[453] In 88% of cases, 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 attempt and/or insulted the victim. Audience anonymity was cited as one of the potential factors contributing to this online baiting behaviour.[454]

6D.69 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 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.[455]

6D.70 A study looking at images posted on Instagram found that the more graphic the self-harm image, the more comments that post seemed to have. While most comments were offering support, and were empathic or neutral in nature, a few comments were hostile.[456]

Reacting to content and re-posting or forwarding content

6D.71 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.[457]

Posting content (text, images, videos)

[450] 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].

[451] 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.13, *PLoS ONE,* 13 (5). [accessed 10 July 2023].

[452] 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].

[453] Videos analysed took place between 2001 and 2017 and included cases from the USA, India, UK, France, Turkey, Canada, Russia, Sweden, Japan and Thailand.

[454] Phillips, J G. and Mann, L., 2019. Suicide baiting in the internet era p.1, *Computers in Human Behaviour,* 92. [accessed 11th July 2023].

[455] Ekō, 2023. Suicide, Incels, and Drugs: How TikTok's deadly algorithm harms kids p.10. [accessed 11 July 2023].

[456] 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].

[457] 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].

6D.72 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.

6D.73 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.[458]

6D.74 A small-scale qualitative study in the UK looking at 18-24-year-olds who had previously had suicidal thoughts, found that the majority of 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/supportive sense of community. [459] However, some noted pessimistic posts as the main factor generating negative effects.

6D.75 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*".[460] 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').[461]

Content exploring

Content tagging

6D.76 The ability to tag content, such as hashtags, are also a risk factor for disseminating suicide or self-harm 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.[462]

6D.77 In other cases, variations of hashtags that are likely to be blocked have been used in an attempt to continue to access the content.

[458] 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].

[459] 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].

[460] 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].

[461] 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].

[462] 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].

User-generated content search filtering

6D.78 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.

6D.79 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.[463]

Hyperlinking

6D.80 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.[464]

6D.81 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.[465]

Recommender systems

Content recommender systems

6D.82 Some evidence suggests that content recommender systems[466] can increase the risk of exposure to self-harm and suicide 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.[467]

6D.83 The coroner's report following the death of Molly Russell, who took her own life following exposure to a large volume of harmful online content, noted how recommender systems can be as a gateway to binge-watching due to the ease of navigation (i.e., scrolling through

[463] 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].

[464] 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].

[465] 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].

[466] 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.

[467] Samaritans and Swansea University, 2022. How social media users experience self-harm and suicide content. p.4. [accessed 10 July 2023]

images and videos). Recommender systems are capable of curating continuous/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.[468] While pieces of content judged in isolation may not be considered illegal, cases such as that of Molly Russell demonstrate the potential cumulative impact and risks of harm amounting from sustained exposure to suicide/self-harm content propagated by recommender algorithms.[469]

6D.84 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".[470]

6D.85 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 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. A separate four accounts were created with a username that indicated a body image-related concern.[471] The researchers found that the 'standard' teen TikTok accounts were recommended self-harm, suicide and eating disorder content 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.[472]

Risk factors: Business models and commercial profiles

6D.86 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[468] The Coroner's Service, 2022. Prevention of Future Deaths. [accessed 28 October 2022].

[469] The Coroner's Service, 2022. Prevention of Future Deaths. [accessed 28 October 2022].

[470] University of Sheffield (Dr Ysabel Gerrard), How we're helping social media companies remove harmful content and protect their users. [accessed 10 January 2023].

[471] Across all accounts, researchers expressed an interest in body image, mental health and eating disorders by watching and liking relevant videos

[472] 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].

6E. Harassment, stalking, threats and abuse offences

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

Summary analysis for harassment, stalking, threats and abuse offences: How harms manifests online, and risk factors

This chapter covers offences relating to online harassment, stalking, threats or abuse which are unwanted behaviours that can cause alarm and distress to other individuals, or put them in fear of violence. 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.** Due to their role in propagating these offences, social media services and online gaming services have been included in the risk profiles.

These offences are also perpetrated on **private messaging services** and **online dating services**.

User base risk factors:

Evidence suggests that anyone can be subjected to these behaviours. However, there is more information about the adverse experiences of **women** on social media. Their exposure to these offences is often more prevalent, severe, and of greater impact compared to men, 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. The general risk factor of **user base demographics** is therefore particularly relevant for the offences listed in this chapter.

Functionalities and recommender system risk factors:

Several functionalities of 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. Due to their role in enabling persistent contact, anonymous user profiles, user profiles and fake user profiles are included in the risk profiles.

In some cases, perpetrators can leverage the **user connections** functionality by connecting with second- and third-degree connections of the victim or survivor, in order 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, which can be leveraged to facilitate harassment and abuse. Individual perpetrators can incite their network to join the abuse of an individual.[473] Due to their use in cases of harassment and stalking, user connections have been included in the risk profiles.

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 contexts and for individuals in the public eye. Abuse and harassment can also occur via **comments on content**. For these reasons, posting or sending location information, reposting or forwarding content, direct messaging and commenting on content have also been included in the risk profiles.

Perpetrators also exploit **user tagging** as a way to harass their victims and survivors by incessantly tagging their usernames in abusive and threatening messages.

Introduction

6E.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

a) content on U2U services that may amount to the harassment, stalking, threats and abuse offences listed under 'Relevant offences' below, and

b) the use of these services for the commission and/or facilitation of these offences (collectively, the ‘risks of harm’).

6E.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.

[473] 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.

6E.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.[474] A case can involve several types of behaviour, hence the grouping of these offences for this chapter.

6E.4 In this chapter:

a) We discuss specific types of threats (such as threats to kill, rape threats, threats of violence) together as 'violent threats'.[475]

b) We use the term 'cyberstalking' in reference to literature that uses this term. While there is no legal definition of 'cyberstalking' the existing harassment and stalking offences cover these behaviours regardless of whether they take place online or offline.

c) 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.

6E.5 Stalking, harassment, threats and abuse offences also occur as part of, or in conjunction with, several other online harms explored in this Register. Key overlaps include chapter 6B: Terrorism offences, chapter 6F: Hate offences, 6G: Controlling or coercive behaviour offences, chapter 6M: Intimate image abuse offences, chapter 6S: Cyberflashing, and chapter 6P: Foreign interference offence.

Relevant offences

6E.6 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to 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.

6E.7 In this chapter, we consider the following offences:

a) Making a threat to kill[476]

b) Behaving in a threatening or abusive manner likely to cause fear or alarm[477]

c) Using threatening, abusive or insulting words or behaviour, with intent to cause fear or provocation of immediate violence[478]

d) 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[479]

e) Using threatening, abusive or insulting words or behaviour, or disorderly behaviour, or displaying any writing, sign or other visible representation which is threatening or

[474] Crown Prosecution Service, 2023. Stalking or Harassment. [accessed 28 September 2023].

[475] 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.

[476] Section 16 of the Offences against the Person Act 1861.

[477] Section 38 of the Criminal Justice and Licensing (Scotland) Act 2010 (asp 13).

[478] Section 4 of the Public Order Act 1986.

[479] Section 4A of the Public Order Act 1986.

abusive, within the hearing or sight of a person likely to be caused harassment, alarm or distress nearby[480]

f) Pursuing a course of conduct which amounts to harassment[481]

g) Pursuing a course of conduct which amounts to stalking[482]

h) Pursuing a course of conduct which puts a person in fear of violence[483]

i) Pursuing a course of conduct amounting to stalking which puts a person in fear of violence, or causes serious alarm or distress[484]

6E.8 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6E.9 The priority offences listed above cover a range of behaviours:

a) **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.

b) **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 chapter 6F: Hate offences for further information). Harassment can also be by two or more people against an individual, or harassment against more than one victim or a group.[485]

c) **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.[486] Examples of behaviours associated with stalking include following a person or contacting or attempting to contact a person.

6E.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.[487] Harassment causing fear of violence, stalking causing fear of violence, and stalking causing serious alarm or distress and having a substantial adverse

[480] Section 5 of the Public Order Act 1986.

[481] 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)).

[482] Section 2A of the Protection from Harassment Act 1997; Section 39 of the Criminal Justice and Licensing (Scotland) Act 2010 (asp 13).

[483] 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)).

[484] Section 4A of the Protection from Harassment Act 1997.

[485] Crown Prosecution Service, 2023. Stalking or Harassment. [accessed 28 September 2023].

[486] Crown Prosecution Service, 2023. Stalking or Harassment. [accessed 28 September 2023].

[487] 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. Source: Suzy Lamplugh Trust. 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].

impact on the victim's day-to-day activities are more serious forms of harassment and stalking.

6E.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.[488]

6E.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 (e.g. sending abusive or threatening messages, images or videos on U2U services).

6E.13 For more details on how services can assess whether content amounts to priority illegal content and more general illegal content, please refer to the **Illegal Content Judgements Guidance** (Volume 5, Chapter 26).

How harassment, stalking, threats and abuse offences manifest online

6E.14 This section is an overview which looks at how harassment, stalking, threats and abuse offences manifest online, and how individuals may be at risk of harm.

6E.15 Reviewing Ofcom and third-party evidence indicates that these offences disproportionately affect certain identity groups – most notably women – alongside other intersecting risk factors; the impact on those individuals can be severe.

6E.16 The UK national institute for data science and artificial intelligence, the Alan Turing Institute, describes the evidence for understanding the presence of online abuse as *"fragmented, incomplete and inadequate*."[489] Crime surveys provide an indication of the number of people affected by online harassment and stalking. The Scottish Crime and Justice Survey asks respondents about experiences of being insulted or harassed online. For 2017/18 it reported that 14% of adults reported experiencing harassing or intimidating behaviour and that, of these, 16% had encountered such behaviour digitally.[490] The 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.[491]

6E.17 Estimates of the presence of online abuse more broadly are limited. Few measure interpersonal online abuse as isolated from abuse towards protected groups (explored in chapter 6F: Hate Offences). However, the available evidence suggests that a substantial portion of the UK population is affected by online abuse. The Oxford Internet Survey in 2019 reported that 27% of people had been exposed to abusive content (including hate

[488] Crown Prosecution Service, 2023. Stalking or Harassment. [accessed 28 September 2023].

[489] The Alan Turing Institute (Vidgen, B., Margetts, H., Harris, A.), 2019. How much online abuse is there? [accessed 28 September 2023].

[490] Scottish Government. 2018. Scottish Crime and Justice Survey 2017/18. [accessed 28 September 2023].

[491] Office for National Statistics, 2022. Stalking: findings from the Crime Survey for England and Wales. [accessed 28 September 2023].

speech).[492] A Pew Research study in 2020 found that 41% of Americans had reported personally experiencing online abuse and harassment.[493]

6E.18 Evidence exclusively measuring the presence of violent threats is also 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 between 2014 and 2020: from 7% to 14% of US adults having experienced it.[494]

6E.19 The Covid-19 pandemic is found to have heightened the risk of stalking perpetrated using online services, and specifically on social media services.[495] 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.[496]

Risks of harm to individuals presented by these offences

6E.20 Harassment, stalking, threats and abuse offences manifest as a wide variety of online behaviours that cause fear, distress and alarm to victims and survivors.

6E.21 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.[497] 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.

6E.22 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.[498] 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.[499] Anyone can be affected by these behaviours, but evidence suggests that

[492] 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].

[493] 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].

[494] Pew Research (Vogels, E.), 2021. The State of Online Harassment. [accessed 28 September 2023].

[495] Titherade & Thomas. 2021. BBC News: Stalking rises during Covid pandemic - police. *BBC News*. 13 July. [accessed 12 July 2023].

[496] Suzy Lamplugh Trust. 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].

[497] Crown Prosecution Service, 2023. Stalking or Harassment. [accessed 28 September 2023].

[498] Refuge. 2021. Unsocial Spaces. [accessed 28 September 2023].

[499] 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. ; Patton, D. U., Pryooz, D., Decker, S., Frey, W. R., Leonard, P. 2019. When Twitter Fingers

women are disproportionately affected.[500] Alongside gender, other demographic factors increase the risks of harm, explored in greater detail in the Risk factors: user base section below.

6E.23 Harms caused by these offences are often severe and varied. 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.[501] A study from 2016 found that 27% of US internet users censor their own online posts for fear of being harassed.[502] 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%).[503]

6E.24 A common impact for victims and survivors of online harassment 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 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:[504]

a) 27% of victims and survivors experienced trouble in a relationship or friendship because of something that was posted about them online;

b) 20% had had to shut down an online account or profile because of online harassment or abuse; and

c) 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.

6E.25 A systematic review of the literature on the experiences of victims and survivors found that cyberstalking can have adverse functional, physiological, and psychological impacts:[505]

a) Functional – including lower professional or academic performance and financial costs.

Turn to Trigger Fingers: a Qualitative Study of Social Media-Related Gang Violence. and Serious Youth Violence. [accessed 28 September 2023].

[500] 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].

[501] Amnesty International, 2017. Social media can be a dangerous place for UK women. [accessed 28 September 2023].

[502] Lenhart, A., Ybarra, M., Zickuhr, K. and Price-Feeney, M., 2016. Online Harassment, digital abuse and cyberstalking in America. [accessed 28 September 2023].

[503] Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S., 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].

[504] Lenhart, A., Ybarra, M., Zickuhr, K. and Price-Feeney, M., 2016. Online Harassment, digital abuse and cyberstalking in America. [accessed 28 September 2023].

[505] Kaura, P., Dhirb, A., Tandone, A., Alzeibyg, E.A. and Abohassanf, 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].

b) Physiological – including changes to eating habits and disrupted sleep patterns.[506]

c) Psychological – including mental and emotional distress, fear, depression, anxiety, withdrawal from social activities.

6E.26 Impact varies by gender. Amnesty international found that more than half of women (55%) who experienced abuse or harassment experienced stress, anxiety or panic attacks following the abuse, while 36% of women said it made them feel that their physical safety was threatened.[507] 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.[508] Fourteen per cent of men found their most recent experience of online harassment 'very' or 'extremely' upsetting, compared to 34% of women.[509] Impacts can also vary for victims and survivors of different ethnicities.[510]

Evidence of risk factors on user-to-user services

6E.27 We consider that the risk factors below are liable to increase the risks of harm relating to harassment, stalking, threats and abuse offences. This is also summarised in the grey box at the start of the chapter.

Risk factors: Service types

6E.28 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

6E.29 Social media services are a risk factor for these offences. According to UNESCO, "*social media companies are the main enablers of online violence against women journalists*", which includes sexual violence and 'gendered profanities'.[511] Stalking, harassment and violent threats can occur on various online services such as online dating services and gaming services but it happens most prolifically on social media services.[512]

[506] 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: C Criado Perez, 2015. Caroline Criado-Perez's speech on cyber-harassment at the Women's Aid conference. *The New Statesmen*, 27 September. [accessed 28 September 2023].

[507] Amnesty International, 2017. Social media can be a dangerous place for UK women. [accessed 28 September 2023].

[508] Lenhart, A., Ybarra, M., Zickuhr, K. and Price-Feeney, M., 2016. Online Harassment, digital abuse and cyberstalking in America. [accessed 28 September 2023].

[509] Pew Research, 2021. The State of Online Harassment. [accessed 28 September 2023].

[510] 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].

[511] 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].

[512] 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.;

Online gaming services

6E.30 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 18-45 had experienced harassment in online multiplayer games, while 71% had experienced severe abuse, including physical threats, stalking, and sustained harassment.[513] 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.[514] The risk can be elevated by livestreaming; gamers in a 2021 study believed that it leads to more bullying behaviour.[515]

6E.31 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*."[516] Strategies used by women to deal with online game-related harassment include leaving online gaming, avoiding playing with strangers, or camouflaging their gender.[517] There is less evidence of male or intersectional experiences.

Online dating services and private messaging services

6E.32 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 women (aged 18-34), with 57% receiving an explicit message, 44% being called an offensive name, and 19% receiving threats to physical harm.[518]

6E.33 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,[519] and another that 48% of female journalists had been harassed by unwanted private messages.[520]

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].

[513] Anti-defamation League, 2021. Hate is No Game: Harassment and Positive Social Experiences in Online Games 2021. [accessed 28 September 2023].

[514] Marwick, A, 2021. Morally Motivated Networked Harassment as Normative Reinforcement. [accessed 28 September 2023].

[515] 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].

[516] 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].

[517] 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].

[518] Online dating users refers to respondents who say they have ever used an online dating site or app (n=2,094). Anderson, M., Vogels, E., Turner, E, 2020. The Virtues and Downsides of Online Dating. [accessed 28 September 2023].

[519] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].

[520] See Risk factors: functionalities and recommender systems section for more information. Source: Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S, 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].

Risk factors: user base

User base demographics

6E.34 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.

6E.35 The data suggests that user-base characteristics: **age, gender, race and ethnicity** could affect the risks of harm to individuals.

6E.36 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 18-34 were more likely than the average internet user to have seen or experienced stalking cyberstalking or harassment in the past four weeks (3% vs 2%).[521] Services with a large number of younger adult users may therefore be disproportionately open to the risks of harm from harassment, stalking, threats and abuse.

6E.37 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, especially forms of harassment that the study classifies as 'severe'.[522]

6E.38 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.

6E.39 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 18-24. 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 assault. [523]

6E.40 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 impacted: 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%).[524]

[521] Comprises waves 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of using online services. [accessed 28 September 2023].

[522] 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].

[523] Amnesty International, 2017. Social media can be a dangerous place for UK women. [accessed 28 September 2023].

[524] Amnesty International, 2017. Social media can be a dangerous place for UK women. [accessed 28 September 2023].

6E.41 Evidence suggests that women in public-facing professions are particularly affected by harassment, threats and abuse. A UNESCO study found that 25% of the women journalists sampled had received threats of physical violence and 18% threats of sexual violence.[525] And 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.[526]

6E.42 The National Crime survey found that 8.3% of women and 3.1% of men reported having experienced cyberstalking specifically.[527] 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.[528]

6E.43 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.[529] Strategies to deal with negative experiences include leaving online gaming, avoiding playing with strangers, or disguising their gender, suggesting that the risks of harm to individuals are higher for players who present as female in these environments.[530]

6E.44 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 25-34 (5.1%), and cyberstalking among women aged 20-24 (3.5%).[531] 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.[532]

6E.45 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 minority ethnic groups, with 3% having seen or experienced 'stalking, cyberstalking or harassing behaviour' in the past four weeks, compared to 1% of white respondents.[533]

[525] Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S., 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].
[526] Inter-Parliamentary Union, 2016. Sexism, harassment and violence against women parliamentarians. [accessed 28 September 2023].
[527] Office for National Statistics, 2022. Stalking: findings from the Crime Survey for England and Wales. [accessed 28 September 2023].
[528] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].
[529] 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].
[530] 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].
[531] Office for National Statistics, 2022. Stalking: findings from the Crime Survey for England and Wales. [accessed 28 September 2023].
[532] Pew Research, 2021. The State of Online Harassment. [accessed 28 September 2023].
[533] Comprises Wave 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of using online services. [accessed 28 September 2023].

6E.46 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.[534] Fifty-four per cent of Black (54%) and 47% of Hispanic online harassment targets said they were harassed due to their race or ethnicity, compared with 17% of white targets.[535]

6E.47 Race and gender are also intersecting risk factors. As Francisco and Felmlee (2021) describe: "online harassment aimed at women of colour emanates from such a hierarchical social system within which women of colour are placed on the lower strata". Black women are significantly more likely to be targets of abuse messages on social media services.[536] The National Crime survey finds stalking to be most prevalent among Black or Black British women (6.5%).[537]

6E.48 OET data suggests that this online harm is experienced more by internet users in certain religious groups, with claimed prevalence among Muslims at 4% compared to just 1% for Christians.[538]

6E.49 Stalking, cyberstalking or intrusive behaviour is higher among those who identify as bisexual. OET data found that while 2% of heterosexuals claimed experience of this harm, bisexuals were more likely to do so, at 4%.[539] 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.[540]

6E.50 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%).[541]

[534] Pew Research, 2021. The State of Online Harassment. [accessed 28 September 2023].

[535] Pew Research, 2021. The State of Online Harassment. [accessed 28 September 2023].

[536] 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(1-3). [accessed 28 September 2023].

[537] Office for National Statistics, 2022. Stalking: findings from the Crime Survey for England and Wales. [accessed 28 September 2023].

[538] Comprises Wave 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of using online services. [accessed 28 September 2023].

[539] Comprises Wave 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of using online services. [accessed 28 September 2023].

[540] FRA EU Agency for Fundamental Rights, 2020. LGBTI Survey Data Explorer. [accessed 28 September 2023].

[541] Lenhart, A., Ybarra, M., Zickuhr, K. and Price-Feeney, M.,2016. Online Harassment, digital abuse and cyberstalking in America. [accessed 28 September 2023].

6E.51 OET data suggests that this online harm is experienced more by internet users with any limiting and impacting conditions (2%), and those with mental health conditions (3%) compared to 1% with no such conditions.[542]

Risk factors: Functionalities and recommender systems

User identification

User profiles

6E.52 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.[543]

6E.53 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 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.[544]

Fake user profiles

6E.54 Moreover, 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.[545] 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.

[542] Comprises Wave 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of using online services. [accessed 28 September 2023].

[543] 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].

[544] U.S Department for Justice (Morgan, R. and Truman, J.), 2022. Stalking Victimization, 2019. [accessed 28 September 2023].

[545] Yardley, E. 2021. Technology-facilitated domestic abuse in political economy: a new theoretical framework, *Violence against women,* 27 (10). [accessed 28 September 2023].

6E.55 Several studies report perpetrators of abuse and harassment creating fake user profiles.[546] 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.[547] 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.[548] 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"*.[549]

Anonymous user profiles

6E.56 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.[550] However, the evidence generally suggests that the ability to create anonymous user profiles also increases the risks of harm explored in this chapter.

6E.57 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 in physical interactions.[551] A 2017 study into the trolling of the McCann family reported that anonymous perpetrators, being unidentifiable as individuals, aligned themselves strongly with group identities and norms in ways that enabled trolling behaviour.[552]

6E.58 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%).[553]

[546] 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].

[547] Refuge, 2022. Marked as Unsafe. [accessed 28 September 2023].

[548] 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].

[549] 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].

[550] McGlynn, C. and Woods, L., 2022. Violence against women and girls (VAWG) Code of Practice. [accessed 28 September 2023].

[551] Suler, J, 2004. The online disinhibition effect. [accessed 28 September 2023].

[552] Synnott, J., Coulias, A. and Loannou, M., 2017. Online trolling the case of Madeleine McCann. [accessed 28 September 2023].

[553] Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S., 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].

6E.59 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.[554] 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.[555]

User networking

User connections

6E.60 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.

6E.61 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'. [556]

6E.62 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.[557]

User tagging

6E.63 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.[558]

[554] 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].

[555] 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].

[556] 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].

[557] 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 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].

[558] 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].

User communication

Livestreaming

6E.64 Several studies suggest that female gamers are at particular risk of abuse and harassment in livestreaming contexts. Online gaming services have been identified as sites of 'normalised harassment', where name-calling or insults are part of the culture.[559] 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.

6E.65 Unbalanced anonymity while livestreaming can create a risk. For example, on gaming services, live-streaming gamers are fully visible to viewers, while viewers can remain fully anonymous. Gamers in a 2021 study believed that this leads to more bullying behaviour.[560]

6E.66 Livestreaming functionalities that encourage the visibility and sharing of streamers' identifiable characteristics create risks of online stalking. For example, even if streamers on gaming services change their username and channel name to escape bullying, perpetrators re-identify victims and survivors based on their video, image or sound, and continue the harassment.[561]

Direct messaging

6E.67 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.[562]

6E.68 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 predominantly direct messaging).[563]

Commenting on content

6E.69 Some services allow users to reply to or comment on posted content. As noted in chapter 6F: Hate offences 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 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 nevertheless calculated to produce alarm or distress, and are oppressive and unreasonable.

[559] Marwick, A, 2021. Morally Motivated Networked Harassment as Normative Reinforcement. [accessed 28 September 2023].

[560] Zhou, Y. and Farzan, Y., 2021. Designing to stop live-streaming cyberbullying. [accessed 28 September 2023].

[561] Zhou, Y. and Farzan, Y., 2021. Designing to stop live-streaming cyberbullying. [accessed 28 September 2023].

[562] Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S., 2020. Online Violence Against Women Journalists. [accessed 28 September 2023].

[563] Suzy Lamplugh Trust, 2021. Unmasking stalking: a changing landscape. [accessed 28 September 2023].

6E.70 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.[564] 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.[565] 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.[566]

Posting content

6E.71 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.[567] 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.[568]

6E.72 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.[569]

Posting or sending location information

6E.73 A 2021 UK study on social media users found that 9.5% of users surveyed reported "*tracking someone through GPS*".[570] 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 tracking might be carried out through spyware devices or apps.[571] Some U2U services host geo-tagging functionalities[572] and this information can be communicated to users.

[564] The Alan Turing Institute (Vidgen, B., Margetts, H. and Harris, A.), 2019. How much online abuse is there? [accessed 28 September 2023].

[565] 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].

[566] YouTube, 2022. YouTube Community Guidelines enforcement – Google Transparency Report. [accessed 25 August 2023].

[567] 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].

[568] 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].

[569] 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].

[570] 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].

[571] *[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].

[572] 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].

6E.74 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 chapter 6G: Controlling or coercive behaviour, the visibility of location information can also be used to track the location of a victim or survivor of previous abuse, including domestic abuse.[573] 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.[574]

Re-posting or forwarding content

6E.75 User-generated content can be decontextualised and reposted, which results in 'context collapse'.[575] 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 a large number of 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.[576]

Recommender systems

Content recommender systems

6E.76 Content recommender systems may amplify the risk of services being used to commit or facilitate harassment, stalking and threats of violence. These types of harmful behaviour may have high engagement rates, which content recommender systems in particular are usually designed to optimise.[577] Recommender systems may therefore reinforce these harmful behaviours by promoting the circulation of harmful or illegal content or connecting users who engage in such illegal activity because they share similar views. However, more technical research is needed to fully understand the role that different recommender systems might play in suggesting specific types of harmful or illegal content.

Risk factors: Business models and commercial profiles

6E.77 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[573] 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].

[574] Waldman, A, 2021. Navigating Privacy on Gay-Oriented Mobile Dating Applications. [accessed 28 September 2023].

[575] 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].

[576] Thrasher, S.W, 2017. Yes there is a free speech crisis. But it's victims are not white men. The *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].

[577] Hateful, harassing content often has high engagement rates, once it evades detection from content moderation systems, it is spread and amplified by platforms' ranking and recommendation algorithms faster than other types of content. Source: Anti-Defamation League response to 2022 Ofcom Call for Evidence: First phase of online safety regulation.

6E. (Continued) Threatening Communications Offence

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

Introduction and relevant offence

6E.78 Section 181 of the Online Safety Act creates the 'threatening communications offence' as a new offence to consider in the Register. 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 and threats chapter. We consider them very likely to manifest in the same way.

6E.79 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:

a) serious injury amounting to grievous bodily harm;

b) rape;

c) assault by penetration; or

d) serious financial loss.

6E.80 The Act also covers the offences of encouraging and assisting, and conspiracy to commit, this offence.

6E.81 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**.

How threatening communications manifest online

6E.82 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 and threat offences chapter. We have therefore made an assumption that it will manifest online in the same or similar ways, although we will keep this under review as the evidence base develops.

6E.83 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.

6E.84 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.[578] [579] 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.[580] Evidence also suggests that the prevalence of threatening communications is high among certain groups, such as women in public roles.[581]

Risks of harm to individuals presented by threatening communications offences online

6E.85 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.

6E.86 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 also refers to this offence in chapter 6G Controlling or coercive behaviour offences and chapter 6K: Sexual exploitation of adults offences.

[578] 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].

[579] 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.

[580] 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].

[581] Amnesty International., 2018. Toxic Twitter – Women's Experiences of violence and abuse on Twitter', Chapter 3 in Online Violence against Women. Source: UNESCO (Posetti, J., Aboulez, N., Bontcheva, K., Harrison, J. and Waisbord, S., 2020. Online violence Against Women Journalists: A Global Snapshot of Incidence and Impacts. [accessed 28 September 2023].

6F. Hate offences

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 can contribute to this risk factor. **Social media services** and **online gaming services** pose a particular risk of hate offences, and have therefore been included in the risk profiles. **Video-sharing services** and **private messaging services** have also been identified as spaces that can be used to commit or facilitate offences related to hate, targeting minorities and other protected groups.

User base risk factors:

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**.

The hate offences discussed in this chapter identify users' **race and ethnicity, religion** and **sexual orientation** as risk factors in their experience of hateful content. But a user's experience of hateful content may also be influenced by their age and gender.

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 the evidence shows that in many instances, hateful content is shared by identifiable users who are unknown to the target. Anonymous user profiles have been included in the risk profiles because of the potential risk they pose for the perpetration of hate offences.

Content recommender systems are another risk factor. Evidence shows that these systems are generally designed to optimise user engagement. In some circumstances this can result in them promoting content which may be harmful in nature. Recommender systems can improve user experiences through offering personalised feeds, but this can also lead to promoting ideas or ideologies that users have already engaged with, which can increase confirmation bias and create ‘filter bubbles’. Content recommender systems have therefore been included in the risk profiles.

Hate offences can be committed via direct responses or **comments on posted content.** This functionality can in particular enable the amplification of hate, known as ‘cybermobbing’ and/or ‘dogpiling’.[582] Commenting on content is a functionality that has been included in the risk profiles. The ability to **livestream** is also a risk factor as hateful content can be broadcast in real time.

Direct messaging can be used to carry out hate offences in a targeted manner and is therefore also included in the risk profiles.

Other functionalities pose risks for hate offences. Evidence suggests that **content tagging** using hashtags is also a risk factor; services have allowed hashtags to include hateful language without detection. **Hyperlinks** allow users to move easily from mainstream to more niche hateful spaces. Usernames on **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.

Business model risk factors:

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. 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

6F.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

a) content on U2U services that may amount to the hate offences listed under ‘Relevant offences’ below; and
b) the use of these services for the commission and/or facilitation of these offences (collectively the ‘risks of harm’).

6F.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

[582] Both terms refer to more than one person directing abusive comments towards an individual.

encountered by an individual, or where content of a particular kind is encountered in combination with content of a different kind.

6F.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

6F.4 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to hate offences, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

6.16 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.[583]
- The publication or distribution of threatening, abusive or insulting written material intending or likely to stir up racial hatred.[584]
- Distributing, showing or playing threatening, abusive or insulting recordings of visual images or sounds intending or likely to stir up racial hatred.[585]
- 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.[586]
- The publication or distribution of threatening written material intending to stir up religious hatred or hatred on the grounds of sexual orientation.[587]
- 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.[588]
- Racially or religiously aggravated harassment and public order offences.[589] This covers various offences relating to the fear or provocation of violence, harassment and stalking, when they are racially or religiously aggravated.

6F.5 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6F.6 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 transgender identity. Crimes can be prosecuted as a hate crime if the offender has demonstrated, or

[583] Section 18 Public Order Act 1986.
[584] Section 19 Public Order Act 1986.
[585] Section 21 Public Order Act 1986.
[586] Section 29B Public Order Act 1986.
[587] Section 29C Public Order Act 1986.
[588] Section 29E of the Public Order Act 1986.
[589] Sections 31 and 32 of the Crime and Disorder Act 1998 and section 50A of the Criminal Law (Consolidation) (Scotland) Act 1995..

been motivated by, hostility on the basis of these characteristics, and this may also have implications for sentencing.[590] For the purpose of 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 chapter 6E: Harassment, stalking, threats and abuse offences.

6F.7 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. Chapter 6E: Harassment, stalking, threats and abuse offences explores threats to kill in more detail. This may involve, or be done alongside, wider threats of violence to the individual.

6F.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** (Volume 5, Chapter 26).

How hate offences manifest online

6F.9 This section is an overview which looks at how hate offences manifest online, and how UK individuals may be at risk of harm.

6F.10 To put these risks of harm into context, Home Office research indicates that instances of online hate crime offences[591] towards a person, based on a protected characteristic[592] in the UK, are widespread. This research shows that 1,605 online hate crimes were recorded in 2017/2018 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).[593] [594]

6F.11 The evidence suggests that a large proportion of people see or experience hateful content online. Ofcom found that during a four-week period, 11% of online adult and child[595] users had seen or experienced content they considered to be hateful, offensive, or discriminatory, and which targeted a group or person based on specific characteristics such as race, religion, disability, sexuality, or gender identity.[596]

[590] The Crown Prosecution Service, n.d. Hate Crime [accessed 7 June 2023].

[591] 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.

[592] 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.

[593] 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.

[594] Home Office, 2018. Hate crime, England and Wales 2017/18. [accessed 2 April 2023].

[595] Children aged 13-17.

[596] Note: Comprises Wave 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of Using Online Services. [accessed 22 September 2023].

6F.12 Ofcom 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.[597] The frequency of hateful content experienced often increased after key national or international events such as a terror attack[598] or large sporting events like Euro 2020.[599] [600] [601]

6F.13 Additionally, research by LGBTQ+[602] charity Stonewall found that one in ten LGBTQ+ people – including one in four trans[603] people – have been the direct target of homophobic, biphobic or transphobic abuse online.[604] Almost half of LGBTQ+ people (45%) have 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

6F.14 Hateful content manifests in several ways online, including posts, memes, and comments on shared content. Ofcom research found that of the respondents who had experienced hateful, offensive or discriminatory conduct online, 47% came across it in comments on or replies to a post, article, or video, while 45% were exposed to the harm by scrolling through a service's feed or 'For You' page.[605]

6F.15 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.[606]

6F.16 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.[607]

6F.17 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.[608] Some users describe experiencing harm immediately after engaging with a one-off

[597] Ofcom, 2023. Qualitative research into the impact of online hate. [accessed March 2023].
[598] Williams, M. and Reya, M 2019. Hatred behind the screens: a report on the rise of online hate speech. [accessed 23 March 2023].
[599]Ofcom, 2023. Qualitative research into the impact of online hate. [accessed March 2023].
[600] 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 Twitter against football players in the 2021-22 Premier League season. [accessed 27 September 2022].
[601] 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].
[602] LGBTQ+ is the acronym recognised by Stonewall.
[603] Term recognised by research authors to describe people whose gender is not the same as, or does not sit comfortably with, the sex they were assigned at birth – Stonewall, no date. List of LGBTQ+ terms. [accessed 8 September 2023]
[604] 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].
[605] Notes: Comprises waves 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of Using Online Services. [accessed 22 September 2023].
[606] See Risk factors: user base section for more information.
[607] Ofcom, 2023. Qualitative research into the impact of online hate. [accessed March 2023].
[608] Ofcom, 2023. Qualitative research into the impact of online hate. [accessed March 2023].

piece of content, while others described experiencing cumulative harm due to repeated exposure to potentially harmful content or interactions.[609]

6F.18 Galop, a UK-based LGBT+[610] 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.[611]

6F.19 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 not for racially motivated crimes. Although the results were inconsistent, they did *"point to the potential for using online behaviour to identify offline risk".*[612]

6F.20 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.[613] 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.

6F.21 While it is challenging to estimate the economic and social cost of hateful content, the UK Government attempted to estimate this for hate crimes that are racially or religiously aggravated to give an indication of the potential scale. The annual cost per crime is estimated to be £16,033 for offences with injury and £6,767 for offences without injury (based on 2021/22 prices). This amounts to a total annual cost of £5.1m in 2021/22 prices, however it was emphasised that this is likely to be underestimated.[614]

Evidence of risk factors on user-to-user services

6F.22 We consider that the risk factors below are liable to increase the risks of harm relating to hate offences. This is also summarised in the grey box at the start of the chapter.

Risk factor: Service types

6F.23 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.

[609] Ofcom, 2022. How people are harmed online: Testing a model from a user perspective. [accessed 2 November 2022].

[610] LGBT+ is the acronym recognised by Galop.

[611] Galop (Hubbard, L.), 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 12 September 2022].

[612] 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].

[613] Council on Foreign Relations, 2018. Hate speech on social media: global comparisons. [accessed 22 May 2023].

[614] The UK government estimated the total annual economic and social cost to be £5.1m in 2021/22 prices (updated by Ofcom to be £5.4m in 2022/23 prices using GDP deflator). However, they emphasised that it is likely to be an underestimate, as it only includes hate offences that are racially or religiously aggravated i.e. it omits other types e.g. offences based on sexual orientation and disability. In addition, prevalence figures used to reach the total cost estimate (718 online hate crimes racially or religiously aggravated) are likely to be underestimated. Source: Department for Digital, Culture, Media & Sport, 2022. Online Safety Bill Impact Assessment. [accessed 22 September 2023].

Social media services

6F.24 Social media services are a risk factor for hate offences targeting minorities and other protected groups.[615] Many social media services publish transparency reports[616] which include metrics for content removed when terms and conditions have been breached, which includes hateful content.[617] The publication of this data shows that social media services recognise that such content is present on their sites.

6F.25 Our evidence shows that some users engage with hateful content on social media services and subsequently form networks of like-minded individuals to proliferate hateful content online.[618] Once communities have been established, evidence indicates that users can be directed by the most active members to less-moderated social media services where hateful ideologies can be discussed more openly.[619]

6F.26 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.[620]

Private messaging services

6F.27 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.[621] 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.[622]

6F.28 Ofcom research revealed how hateful content can be shared via private messaging services: a 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

[615] See section above on Risks of harm to individuals presented by the hate offences for more information.

[616] 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].

[617] Further information on metrics can be found on individual platforms' published transparency reports.

[618] 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].

[619] 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].

[620] Reed, A, Whittaker, J, Votta, F. and Looney, S., 2019. Radical filter bubbles: social media personalisation algorithms and extremist content. [accessed 19 February 2023].

[621] 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].

[622]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].

'raiders' whose usernames and profile pictures contained racist slurs and imagery. The individual felt targeted due to their LGBTQ+ status.[623]

Online gaming services

6F.29 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"*.[624] 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*".[625] This research also found "*antisemitism, racism and homophobia on platforms [where] users stream and chat about games*".[626]

Services enabling online community building

6F.30 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 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*".[627] This study provides anecdotal evidence in relation to how potential perpetrators view online communities as an opportunity to spread and disseminate hateful content.

6.17

Risk factors: User base

User base size

6F.31 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.[628]

6F.32 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

[623] Ofcom, 2022. How people are harmed online: Testing a model from a user perspective. [accessed 2 November 2022].
[624] 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].
[625] Miller, C. and Silva, S, 2021. Extremists using video-game chats to spread hate, *BBC,* 23 September. [accessed 22 September 2023].
[626] Miller, C. and Silva, S, 2021. Extremists using video-game chats to spread hate, *BBC,* 23 September. [accessed 22 September 2023].
[627] 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].
[628] 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].

hateful content more openly.[629] It is possible that these less-moderated services have smaller user base sizes.

User base demographics

6F.33 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.

6F.34 Data suggests that user base characteristics including **age**, **sexual orientation, race and ethnicity, and religion** affect the risks of harm to individuals.

6F.35 Users' likelihood of being exposed to hateful content online appears to be correlated with their age. Ofcom research found that on average, 11% 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, e.g. hate speech"* in the four weeks leading up to the study.[630] Eighteen per cent of younger users aged 18-24 reported seeing such content, compared to 13% of those aged 35-44 and 9% aged 45-54.[631] The Oxford Internet Survey of internet use in Great Britain[632] also found that younger people are more likely to experience abuse online; 41.2% of 18-30-year-olds had seen cruel/hateful content online, compared with 7.4% of over-75s.[633]

6F.36 The ethnicity of a user also affects how likely it is that they will be exposed to hateful content. Ofcom research found that of the 11% of UK internet users who had seen or experienced *hateful, offensive, or discriminatory content* online in the past four weeks, 16% defined themselves as any mixed ethnicity, compared to 10% who defined themselves as white.[634]

6F.37 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.[635]

6F.38 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 to hateful content. Ofcom research found that of the 11% of UK internet users who had seen or experienced hateful, offensive, or discriminatory content online in the past four weeks, Muslim internet users were more likely to report having seen or experienced such content (16%) than Christian users (8%) or respondents with no religion (13%).[636]

[629] 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].

[630] Note: Comprises waves 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of Using Online Services. [accessed 22 September 2023].

[631] 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].

[632] Multi-stage national probability sample of 2,000 people.

[633] 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].

[634] Note: Comprises waves 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of Using Online Services. Online Experiences Tracker 2021-2022. [accessed 22 September 2023].

[635] 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].

[636] Note: Comprises waves 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of Using Online Services. Online Experiences Tracker 2021-2022. [accessed 22 September 2023].

6F.39 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 (21%) and bisexual UK internet users (25%) were more likely to report having seen or experienced such content than heterosexual users (10%).[637]

6F.40 Galop's Online Hate Crime Report surveyed over 1,100 LGBT+ people and found that 60% of respondents had experienced anti-LGBT+ abuse online.[638]

6F.41 The Anti-Defamation League (ADL) published research in 2020 into online hate and harassment.[639] 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 religion-based harassment (from 11% to 22%) while race-based harassment had increased from 15% to 25%.[640]

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles

6F.42 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.

6F.43 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.[641] 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.

6F.44 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,[642] and suggested that weak forms of identity (i.e. anonymity) fuels more hate on online social media.[643]

[637] Note: Comprises Wave 1 and 2 combined data set. Source: Ofcom, 2023. Experiences of Using Online Services. Online Experiences Tracker 2021-2022. [accessed 22 September 2023].

[638] Galop, 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 12 September 2022].

[639] Nationally representative survey of Americans, 1,974 respondents.

[640] ADL, 2020. Online hate and harassment report: The American experience 2020.[accessed 20 October 2022].

[641] Galop, 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 12 September 2022].

[642] 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].

[643] 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].

6F.45 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.[644] Anonymity can enable users to express ideas or criticisms about people in power without risking attribution.[645] 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*".[646] Therefore, while anonymity online may give rise to some risks, it also confers some important benefits.

User profiles

6F.46 Usernames on user profiles can be used to stir up hatred against groups with protected characteristics, by intentionally publishing or distributing threatening material. Evidence suggests that usernames have been used as a tool by users to spread racial slurs. [647] 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.[648] Users can often easily find the banned user via their username, even after the original account has been banned.[649]

User communications

Livestreaming

6F.47 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 chapter 6B: Terrorism offences, chapter 6C: Child sexual exploitation and abuse (CSEA) offences, and chapter 6L: Extreme pornography offence, livestreaming can be used to broadcast illegal content in real time.

6F.48 We expect livestreams can also be used to share hateful content. Livestreams can be used to broadcast hateful content with large audiences of users and their ephemeral nature makes moderation challenging.[650] Moreover, some video-sharing services allow users to combine user-generated content with existing content, which can be used to respond to posts in a hateful way.[651] The ephemeral nature of livestreaming means that the content is less likely to be archived[652] and may not be moderated in real time. The risk of harm presented by livestreaming is increased when paired with screen recording functionality, as the

[644] E-Safety Commissioner, n.d. Anonymity and identiy shielding [accessed 22 May 2023].

[645] E-Safety Commissioner, n.d. Anonymity and identiy shielding [accessed 22 May 2023].

[646] E-Safety Commissioner, n.d. Anonymity and identiy shielding [accessed 22 May 2023].

[647] Institute for Strategic Dialogue (O'Conner, C.), 2021. Hatescape: An in-depth analysis of extremism and hate speech on TikTok. [accessed 9 March 2023].

[648] Institute for Strategic Dialogue (O'Conner, C.), 2021. Hatescape: An in-depth analysis of extremism and hate speech on TikTok. [accessed 9 March 2023].

[649] Institute for Strategic Dialogue (O'Conner, C.), 2021. Hatescape: An in-depth analysis of extremism and hate speech on TikTok. [accessed 9 March 2023].

[650] Zhou, Y. and Farzan, R, 2021. Designing to stop live streaming cyberbullying: A case study of Twitch live streaming platform. [accessed 22 September 2023].

[651] Institute for Strategic Dialogue (O'Conner, C.), 2021. Hatescape: An in-depth analysis of extremism and hate speech on TikTok. [accessed 9 March 2023].

[652] Zhou, Y. and Farzan, R, 2021. Designing to stop live streaming cyberbullying: A case study of Twitch live streaming platform. [accessed 15 February 2023].

subsequent recording and dissemination of potentially hateful livestreamed footage can increase content virality.[653]

Direct messaging

6F.49 As discussed in chapter 6E: Harassment, stalking, threats and abuse offences, 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.[654]

Commenting on content

6F.50 Some services allow users to reply to or comment on posted content. Direct responses or comments on posted content may in some cases amount to illegal content. Sometimes hateful content sent to an individual through a comment functionality can be amplified by the scale of comments that the individual receives. Ofcom research on Twitter footballer abuse suggests that users may send just one abusive comment to an individual,[655] but sometimes that targeted individual can receive comments from a lot of 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’.[656] [657]

Reacting to content

6F.51 Users’ reaction to content can enable the building of networks of like-minded individuals, where hateful content can be shared among a receptive audience. Research examining the ‘Stop Islam’ hashtag on Twitter, 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.[658]

Posting content

6F.52 The ability to post content on services can allow the easy dissemination of content, increasing the risk of exposure of users to controversial or emotive posts. This can also make it easier to disseminate hateful content.

[653] Ofcom, 2022. The Buffalo Attack: Implications for Online Safety. [accessed 18 January 2023].
[654] Ofcom, 2023, Qualitative research into the impact of online hate. [accessed 22 September 2023].
[655] Ofcom’s research into Twitter 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 Twitter against football players in the 2021-22 Premier League season. [Accessed 27 September 2022].
[656] Galop, 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 12 September 2022].
[657] Both terms refer to more than one person directing abusive comments towards an individual.
[658] Poole, R, Giraud, E, Quincey E, 2021. Tactical interventions in online hate speech: The case of #stopIslam, *New Media and Society,* 23(6). [accessed 12 Jan 2023].

6F.53 Research also shows that posts featuring hashtags and containing antisemitic comments can reach large audiences on these services.[659]

6F.54 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 Twitter found that six months into the Covid-19 pandemic, hateful content was largely being produced by bots.[660] 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-generated content exploring

Hyperlinking

6F.55 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[661] and other hyperlinks, audiences on a video-streaming service can easily move from mainstream to hateful content.[662]

Content tagging

6F.56 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 services will promote certain trending topics to wider audiences.[663] 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.[664]

6F.57 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.[665] A high-profile example

[659] CCDH, 2021. Failure to protect: How tech giants fail to act on user reports of antisemitism. [accessed 20 September 2023].

[660] 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].

[661] 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'.

[662] Data & Society (Lewis, R.), 2018. Alternative influence: broadcasting the reactionary right on Youtube. [accessed 20 February 2023].

[663] Institute for Strategic Dialogue, (O'Conner, C.), 2021. Hatescape: An in-depth analysis of extremism and hate speech on TikTok. [accessed 9 March 2023].

[664] This relates to services which employ trending functionalities.

[665] Antisemitism Policy Trust, 2021. Instagram: Bad Influence. [accessed 20 December 2022].

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.[666] 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

6F.58 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.[667] 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.[668]

Recommender systems

Content recommender systems

6F.59 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 recommender system may then promote this content, which could be divisive, untrue, or incendiary.[669]

6F.60 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.

6F.61 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.[670] This suggests that when users interact with

[666] BBC News, 2016. Man jailed for harassing Labour MP Luciana Berger. *BBC News*, 8 December. [accessed 22 September 2023].

[667] Institute for Strategic Dialogue (O'Connor, C.), 2021. Hatescape: An in-depth analysis of extremism and hate speech on TikTok. [accessed 9 March 2023].

[668] CST, 2020. Hate fuel: The hidden online world fuelling far right terror. 11 June. [accessed 7 June 2023].

[669] Munn, L, 2020. Angry by design: Toxic communication and technical architectures. *Humanities and Social Sciences communications*. 7 (53) [accessed 3 May 2023].

[670] Fringe content coded as radical content without justification of violence, may also include profanity laden nicknames that go beyond political discourse, or historical revisionism.

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.[671]

6F.62 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.[672]

Risk factors: Business models and commercial profile

Revenue models

Advertising-based revenue model

6F.63 Services 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, services which rely on advertising revenue models have a financial incentive to promote content that drives user engagement.

6F.64 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: toxic communication and technical architectures*[673] 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 drive up both the number of users and user time spent.

6F.65 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 key 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"*,[674] and *"to break the economic link between advertising and the harmful content".*[675]

Subscription-based revenue models

6F.66 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.[676]

[671] Munn, L, 2020. Angry by design: Toxic communication and technical architectures. *Humanities and Social Sciences communications*. 7 (53) [accessed 3 May 2023].

[672] Munn, L, 2020. Angry by design: Toxic communication and technical architectures. *Humanities and Social Sciences communications*. 7 (53) [accessed 3 May 2023].

[673] Munn, L, 2020. Angry by design: Toxic communication and technical architectures. *Humanities and Social Sciences communications*. 7 (53) [accessed 3 May 2023].

[674] World Federation of Advertisers, 2020. Marketing leaders take action on harmful online content. [accessed 22 September 2023].

[675] Conscious Advertising Network, n.d. About Us. [accessed 22 September 2023].

[676] Stanford (Thiel, D. and Mcain, M.), 2022. Gabufacturing dissent: An in-depth analysis of Gab. [accessed 3 May 2022].

6G. 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. 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 chapter 6E: Harassment, stalking, threats and abuse, and chapter 6M: Intimate image abuse 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** also enable perpetrators to be a constant presence in the lives of victims and survivors. Because of their ability to facilitate coercive control, social media services and private messaging services are included in the risk profiles.

***Online dating services** and **online adult 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. **Young women,** as well as women from **minority ethnic and racial backgrounds,** appear to be most at risk. Research indicates that 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, as well as 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. Fake user profiles and direct messaging have therefore also been included in the risk profiles.

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. Due to their role in enabling CCB, fake user profiles, user connections and posting or sharing location information are included in the risk profiles.

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. Posting content has therefore also been included in the risk profiles.

Introduction

6G.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the CCB 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').

6G.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.

6G.3 This chapter uses the term CCB primarily to reflect the offence listed in the Online Safety Act (the Act). CCB specifically includes engendering psychological fear as a form of abuse.[677]

6G.4 A note on terminology – some evidence cited in this chapter explores 'technology-facilitated domestic abuse' (shortened to 'tech abuse' in some reporting). Given that technology-facilitated domestic abuse engenders psychological as opposed to physical harm, we consider these studies to fall in scope of the legal definition of CCB. The language of coercive control is similarly used by scholars in the field 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

[677] Stark, E., 2009. Coercive control: the entrapment of women in personal Life. [accessed 21 September 2023].

inequality.[678] However, to stay aligned with the evidence, we use the terminology from the research when citing studies in this chapter.

6G.5 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, the vast majority of 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.[679]

6G.6 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

6G.7 The Act requires Ofcom to consider the risks of harm connected with specific offences. With regard to 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.[680]

6G.8 CCB occurs where the victim 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, 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.

6G.9 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.[681]

6G.10 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.

6G.11 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of the offence.

[678] For example, see source: 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., Woodlock, D., 2022. Digital coercive control: Insights from two landmark domestic abuse studies. *The British Journal of Criminology*, 59(3). [accessed 21 September 2023].

[679] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[680] Section 76 of the Serious Crime Act 2015.

[681] Crown Prosecution Service, 2023. Legal Guidance: Controlling or Coercive behaviour in an intimate or family relationship. [accessed 5 September 2023].

6G.12 For more details on how services can assess whether content amounts to priority illegal content, and more generally illegal content, please refer to the **Illegal Content Judgements Guidance** (Volume 5, Chapter 26).

How controlling or coercive behaviour manifests online

6G.13 This section provides an overview on how CCB manifests online, and how individuals may be at risk of harm.

6G.14 To put the risks of harm into context, a survey by Women's Aid[682] of victims and survivors of domestic abuse in 2013 found that 45% had experienced abuse online during their relationship.[683] Research by Refuge[684] conducted in 2022 found that 82% of victims and survivors of tech abuse had experienced harassment or stalking on social media, 41% had experienced threats of violence and 29% had experienced intimate image abuse.[685] However, getting up-to-date measurement of CCB, including the specific prevalence of online CCB, is challenging. It is not only a relatively new offence, but also overlaps with other offences (see in 6G.10).

6G.15 Low incidences of reporting further complicate the measurement of CCB; like related offences, CCB is consistently under-reported.[686] 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.[687] Evidence also suggests that women may not identify as victims and survivors if asked directly; responses are higher if women are given specific examples of abuse to relate to.[688]

Risks of harm to individuals presented by CCB

6G.16 U2U 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.[689]

[682] Women's Aid is a charity working to end domestic abuse against women and children.

[683] Women's Aid, (Laxton, C.), 2014. Virtual World, Real Fear: Women's Aid Report into Online Abuse, Harassment and Stalking. [accessed 23 September 2023].

[684] Refuge is a charity providing specialist support for women and children experiencing domestic violence.

[685] Based on 17 qualitative interviews with women supported by Rufuge. Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[686] 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. Source: National Police Chiefs Council, 2023. Violence Against Women and Girls: Strategic Threat Risk Assessment 2023. [accessed 21 September 2023].

[687] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[688] Refuge reports 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. Source: Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].; Similarly, an Australian study found that many victims and survivors of intimate partner stalking do not identify stalking behaviour as such. Source: Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

[689] Harris, B., Woodlock, D., 2022. Digital coercive control: Insights from two landmark domestic abuse studies. *The British Journal of Criminology,* 59(3). [accessed 21 September 2023].

6G.17 U2U services have shifted the landscape of domestic abuse by better enabling perpetrators to coerce and control victims and survivors at a distance.[690] They can enable a sense of perpetrator omnipresence, due to the wide scope of the technology, with victims and survivors feeling they cannot escape the perpetrator's abuse.[691] Nearly all victims and survivors of online CCB report experiencing harmful outcomes of varying forms and severity.

6G.18 Some behaviours replicate or extend the dynamics of offline coercion; for example, sending abusive messages. Others present more online-specific methods of abuse; sharing or threatening to share intimate images online without consent, and hacking victims and survivors' accounts are now common methods in cases of CCB, and enable additional forms of coercion and control.[692] 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,[693] as well as access and potentially share intimate images from that individual's account.[694] Intimate image abuse in the context of CCB will be discussed in more detail in paragraph 6G.65.

6G.19 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.[695]

6G.20 Women are disproportionately affected by CCB. It is most commonly perpetrated by men with the aim of controlling or coercing their former or current partner.[696] While this refers to domestic abuse more broadly, this balance is likely to be reflected for CCB specifically.[697]

6G.21 LGBTQ+ populations are also likely to suffer more severe outcomes from partner abuse than their heterosexual equivalents. A study in 2017 found that LGBT+ victims and survivors are almost twice as likely to have attempted suicide, more than twice as likely to have self-

[690] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[691] Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

[692] Refuge found that 47% of victims and survivors reported that someone had access to their social media account against their wishes. Source: Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[693] Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[694] An Australian study reports a perpetrator hacking a victim's and survivor's account, before sharing intimate images with male contacts. Refuge reports 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. Source: Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023]; Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[695] The most common forms of CCB online identified by Refuge are: 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 'doxing.' Source: Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[696] In the year ending March 2022, the victim was female in 74.1% of domestic abuse-related crimes. Source: Office of National Statistics, (Elkin, M.), 2022. Office of National Statistics, 25 November 2022. Domestic abuse victim characteristics, England and Wales: year ending March 2022. [accessed 8 September 2023].

[697] 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.

harmed, more likely to face abuse from multiple perpetrators and twice as likely to have experienced historic abuse from a family member.[698]

6G.22 Harm is also not restricted to victims and survivors. Perpetrators of online CCB can involve victims' and survivors' family and friends to extend their abuse and increase omnipresence.[699] Refuge found that half of the women surveyed (50%)[700] said that their family or friends had been targeted as part of the online abuse, including 12% who reported that their children had been targeted by their partner or former partner.[701]

6G.23 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.

6G.24 Psychological harm is almost universal; the evidence consistently reveals emotional turmoil, life complications, and helplessness among victims and survivors.[702] 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.[703] More than one in three women felt anxious and stressed (37% and 36%), one in five felt ashamed and isolated (21% and 19%), while one in ten felt suicidal as a result of the abuse.[704]

6G.25 Online CCB may indicate a risk of physical violence and loss of life,[705] although there is limited evidence of this to date.[706] 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*".[707]

6G.26 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.[708] Fifteen per cent felt their physical safety was more at risk, 5% felt more at risk of 'honour'-based violence,[709] and 12% felt afraid to leave the house

[698] Office of National Statistics, (Bradley, A., Potter, A.), 2018. Women most at risk of experiencing partner abuse in England and Wales: years ending March 2015 to 2017. [accessed 21 September 2023].

[699] Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

[700] 50% = 579 Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[701] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[702] 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].

[703] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[704] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[705] Practitioners in Australia have identified technology-facilitated stalking and abuse as indicators that women and children are at risk of domestic homicide. Source: Woodlock, D., McKenzie, M., Western, D., Harris, B., 2020. Technology as a Weapon in Domestic abuse: Responding to Digital Coercive Control. *Australian Social Work,* 73 (3). [accessed 21 September 2023].

[706] 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].

[707] All-Party Parliamentary Group on Domestic Violence. Source: Women's Aid, 2017. Tacking domestic abuse in the digital age. [accessed 21 September 2023].

[708] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[709] Honour-based abuse is a crime or incident committed to protect or defend the 'honour' of a family or community. Source: Metropolitan Police, n.d. What is honour-based abuse?. [accessed 5 September 2023].

because of online abuse.[710] Fear of physical harm is likely to contribute to the near-universal psychological harm.

6G.27 Some victims and survivors are left feeling uncomfortable online.[711] 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.[712]

6G.28 The risks of harm are increased by the length of time online CCB lasts, with research showing that CCB perpetrated online often worsens over time.[713] Research also found that risk of online CCB appears to increase once victims and survivors have separated from their partners, known as 'post-separation'. Relationship breakdown can be a common trigger;[714] 66% of women who experienced tech abuse from an intimate partner said they were an ex-partner at the time of the abuse, and 18% said the perpetrator was a partner at the time of the abuse.[715]

6G.29 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, however, likely to be applicable. More information on exploitation can be found in chapter 6K: Sexual exploitation of adults offences chapter.

Evidence of risk factors on user-to-user services

6G.30 We consider that the risk factors below are liable to increase the risks of harm relating to CCB. This is also summarised in the grey box at the start of the chapter.

Risk factor: Service types

6G.31 Research indicates that social media services can be used to commit or facilitate CCB. Messaging services, adult services, and dating services may also be risk factors.

Social media services

6G.32 There is strong evidence that a significant proportion of online CCB takes place on social media services. Research by Refuge shows that 36% of UK women have experienced abuse

[710] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[711] Refuge found 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. Source: Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[712] 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."* Source: 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].

[713] Almost half the victims and survivors responding to a survey by Refuge (48%) said that the abuse they experienced on social media got worse over time. Fifteen per cent said the abuse worsened when they reported the perpetrator or took action to mitigate the abuse, such as blocking the perpetrator online. Source: Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[714] Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

[715] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

on social media or other online services.[716] It also shows that 60% of women who reported experiencing abuse on social media services also reported emotional abuse, and 21% experienced coercive control.[717]

6G.33 Social media services can be used in 'proxy stalking', a mechanism that allows perpetrators to monitor and contact victims through other people.[718] 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 some social media services.[719]

Messaging services and dating services

6G.34 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 of victims.[720] Direct messaging is central to messaging services, as are many social media services and dating services.

6G.35 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.[721]

Adult services

6G.36 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.[722] Adult services host significant amounts of intimate image abuse; this and other services which are known to facilitate intimate image abuse are discussed in chapter 6M: Intimate image abuse offences.

Risk factors: User base

User base size

6G.37 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.

[716] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].
[717] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].
[718] An Australian study found multiple cases of perpetrators using the social media pages of shared friends, family or even their children for monitoring purposes. Source: Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].
[719] 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. Source: 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].
[720] See Risk factors: functionalities and recommender systems section for more information.
[721] 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].
[722] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

6G.38 Meta-owned Facebook (including Messenger) and WhatsApp reached 92% and 87% of UK online adults respectively in January 2023.[723] 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.[724] TikTok and Snapchat reached 42% and 21% of UK online adults respectively in January 2023.[725] These services are less commonly used in abuse cases, with 20% of victims and survivors reporting abuse via Snapchat, and 13% on TikTok.[726] 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

6G.39 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.

6G.40 While demographic information can be limited due to the sensitivity and importance of anonymity in CCB cases,[727] our evidence suggests that users' **age, gender, race and ethnicity, sexual orientation, gender identity,** as well as **socio-economic status** could lead to an increased risk of harm to individuals.

6G.41 Age can be a relevant risk factor in cases of domestic abuse. ONS data identifies women aged 16-24 years old as most at risk of partner abuse, with risk decreasing for older cohorts.[728] Evidence suggests that abuse perpetrated on social media is particularly common among young women. Refuge reports a third (30%) of women aged 16-19 having experienced CCB in a relationship, rising to over half (51%) when presented with a list of potentially controlling or coercive behaviours.[729] Twenty-six per cent of young women report having their social media accounts monitored by a partner or former partner, making it one of the most commonly experienced forms of coercive control among young women.[730]

6G.42 Evidence into CCB specifically is limited, although there is a wealth of evidence indicating that domestic abuse more broadly disproportionately affects women. Eighty-three per cent of high-frequency victims and survivors (having experienced more than ten crimes) are women.[731]

[723] Ipsos, Ipsos iris online audience measurement service, Facebook & Messenger and WhatsApp, January 2023, age: 18+, UK.

[724] Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[725] Ipsos, Ipsos iris online audience measurement service, TikTok and Snapchat, January 2023, age: 18+, UK.

[726] Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[727] 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. Source: Grimani, A., Gavine, A., Moncur, W., 2022. An Evidence Synthesis of Covert Online Strategies Regarding Intimate Partner Violence, *Trauma, Violence and Abuse*, 23(2). [accessed 21 September 2023].

[728] 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. Women most at risk of experiencing partner abuse in England and Wales: years ending March 2015 to 2017. [accessed 21 September 2023].

[729] Refuge, 2021. Marked as Unsafe. [accessed 21 September 2023].

[730] Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[731] 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].

6G.43 Women were almost three times more likely than men to experience sexual abuse from the person who abused them online.[732] This suggests that women are more likely to experience online CCB as part of a continuum of abuse.

6G.44 There is also evidence to suggest that CCB is disproportionately perpetrated by men. 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.[733] 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.[734]

6G.45 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. [735] Researchers identify that although the proportions are very low, the numbers of men affected are still substantial (and include gay, bisexual and/or trans men as well as heterosexual cisgender men), with 79,473 men experiencing 219,118 cases of domestic abuse.[736]

6G.46 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 / 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/Asian British women.[737]

6G.47 ONS data from 2017 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%).[738]

6G.48 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.[739] Although online

[732] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[733] 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].

[734] 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].

[735] 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]; Donovan, C., 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].

[736] Donovan, C., 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].

[737] Office for National Statistics, 2018. Women most at risk of experiencing partner abuse in England and Wales: years ending March 2015 to 2017. [accessed 21 September 2023].

[738] Office for National Statistics, 2018. Women most at risk of experiencing partner abuse in England and Wales: years ending March 2015 to 2017. [accessed 21 September 2023].

[739] 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. Source: Galop, n.d. Domestic Abuse. [accessed 5 September 2023].

CCB is not measured specifically, both online abuse and intimate partner abuse are higher among LGBTQ+ populations. LBTQ+ women are much more likely to have experienced online abuse than women who do not identify as LBTQ+. 75% of LBTQ+ female survey respondents said they had experienced online abuse, compared to 33% of non-LBTQ+ women.[740] The ONS found that bisexual women were nearly twice as likely to have experienced partner abuse in the past 12 months as heterosexual women (10.9% vs 6.0%).[741]

6G.49 Evidence linking online CCB and gender identity is also lacking. However, evidence of intimate partner violence more broadly indicates that not being cisgender[742] 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.[743]

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

6G.50 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.

6G.51 Analysis of media reports found that 'fake' user profiles are a common CCB tactic.[744] 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 that represent fictitious people or real people known to victims and survivors.[745] Refuge reports that 29% of victims and survivors have been impersonated.[746]

6G.52 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.[747] 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.[748] More detail on harassment through fake user profiles can be found in chapter 6E: Harassment, stalking, threats and abuse offences.

[740] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[741] Office for National Statistics, 2018. Women most at risk of experiencing partner abuse in England and Wales: years ending March 2015 to 2017. [accessed 21 September 2023].

[742] A person who has a gender identity that matches their sex assigned at birth.

[743] National Center for Transgender Equality, 2015. US Transgender Survey. [accessed 8 September 2023]

[744] 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].

[745] 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. Source: 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].

[746] Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[747] Refuge response to Ofcom 2022 Call for Evidence: First phase of online safety regulation, full quote: "*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]*".

[748] Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

6G.53 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% of victims and survivors reported being contacted repeatedly by an account that they suspected to be fake.[749] Limited recourse exacerbates this issue, with Refuge reporting that services almost never take down suspected fake accounts and their associated fake user profiles.

6G.54 As described in the 6E: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.[750]

User networking

User connections

6G.55 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 seen everything, which is a key 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.[751]

6G.56 In addition to using their own pre-existing networks comprising friends and family,[752] studies report that perpetrators are also using victims' and survivors' networks to intimidate, harass and humiliate women, or challenge women's accounts of abuse.[753] Victims' and survivors' networks are often not aware of how they are being used.

6G.57 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.[754] 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.[755]

[749] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[750] 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].

[751] Melton, H, 2007. Stalking in the context of intimate partner abuse: In the victims' words. *Feminist Crimonolgy* 2(4). [accessed 21 September 2023].

[752] 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. Source: Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[753] 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., Moncur, W., 2022. An Evidence Synthesis of Covert Online Strategies Regarding Intimate Partner Violence. *Trauma, Violence and Abuse,* 23(2). [accessed 21 September 2023].

[754] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[755] Woodlock, D. 2017. The Abuse of Technology in Domestic Abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

User communications

Direct messaging

6G.58 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.

6G.59 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.[756] [757] 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".*[758]

6G.60 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.[759]

Posting content (text, images)

6G.61 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.

6G.62 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 ('doxing'), negative information, intimate images, and threatening words or images.

6G.63 The ability to post content, combined with user networks, facilitates 'doxing.' This describes sharing identifying information about a particular individual online with intent to cause harm or distress. Doxing 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 doxing.[760]

[756] 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].

[757] 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. Source: 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].

[758] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[759] 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].

[760] The Refuge study also provides qualitative examples of doxing 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. Source: Refuge. 2022. Marked as Unsafe. [accessed 21 September 2023].

6G.65 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.[761] An Australian study reported that 33% of victims and survivors have had negative information about them posted on social media services.[762]

6G.66 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 chapter 6M: Intimate image abuse offences) 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.[763] 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.[764] Of the 376 prosecutions for intimate image abuse offences recorded in the year ending March 2019, 83% were flagged as being domestic abuse-related.[765] Other studies provide case studies of intimate image abuse in the context of CCB.[766]

Posting or sending location information, user groups, user events, user tagging

6G.67 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% of perpetrators reported 'tracking someone through GPS'.[767] This section draws on evidence from chapter 6E: Harassment, stalking, threats and abuse offences. More detail on cyberstalking can be found in this chapter.

6G.68 A recent survey by Refuge found that 41% of victims and survivors had experienced location tracking.[768] 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

[761] Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

[762] 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. Source: Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

[763] Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

[764] Sharatt, E., 2019. Intimate image abuse in adults and under 18s. [accessed 21 September 2023].

[765] Office of National Statistics, (Elkin, M.), 2019. Office of National Statistics, 25 November 2019. [accessed 8 September 2023] Domestic abuse and the criminal justice system, England and Wales: November 2019. [accessed 8 September 2023]

[766] 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. Source: Sharratt, E, 2021. Intimate Image Abuse: A consultation paper. [accessed 21 September 2023].

[767] Gunn, R., Tzani, C., Ioannou, M., Synnott, J., Fumagalli, A., 2021. Cyberstalking among social media users: Perceptions, prevalence and characteristics. [accessed 21 September 2023].

[768] Refuge, 2022. Marked as Unsafe. [accessed 21 September 2023].

visible to potential perpetrators.[769] Perpetrators can use geolocation tracking (for example, attached to status updates) to see where their partners and former partners are.[770]

6G.69 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.[771]

Content editing

Editing visual media

6G.70 The editing of visual media such as images and videos to create deepfakes,[772] 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.[773] While no specific evidence exists for this, threatening to create deepfakes may also be present in some cases of CCB.

Recommender systems

Content and network recommender systems

6G.71 Content recommender systems are commonly designed to personalise content, and users who positively engage (e.g., 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 partner's accounts/ stalk partners"*.[774] 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

6G.72 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[769] Woodlock, D. 2017. The Abuse of Technology in Domestic abuse and Stalking. *Violence against women,* 23(5). [accessed 21 September 2023].

[770] 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].

[771] Chaulk, K., Jones, T. 2011., Online Obsessive Relational Intrusion: Further Concerns About Facebook. *Journal of Family Violence,* 26. [accessed 21 September 2023].

[772] 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.

[773] Refuge, 2021. Unsocial Spaces. [accessed 21 September 2023].

[774] 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].

6H. Drugs and psychoactive substances offences

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

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 risks of harm to individuals from drugs and psychoactive substances is well-documented in terms of both physical and mental health. Its impact can be felt on society through negative effects on health services as well as criminality levels. Although there is limited evidence on the online supply of drugs and psychoactive substances, the hidden nature of this illegal activity will contribute to the risks of harm to individuals; there is evidence to show that people see it and are aware of it, in both passive scrolling and in more targeted messages.

Service type risk factors:

Based on the evidence available, **social media services** and **messaging services** appear to be the service types that are most at risk of being exploited by perpetrators buying or offering to supply drugs and other psychoactive substances. As a result, these types of services are included in the risk profiles.

Some **video-sharing services** are also used to access drugs.

User base risk factors:

Gender is not necessarily a defining factor for users who sell and purchase drugs and psychoactive substances online. However, men tended to see more hard drugs promoted online than women. Under-18s tend to receive more content promoting drugs on certain social media services than over-18s.

Functionalities and recommender systems risk factors:

The ability to **post content,** in particular images and emojis, as well as **post goods and services for sale**, 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. **User-generated content searching** is a popular way for users to find illicit drugs and psychoactive substances. 'Menus' or images depicting the products for sale can be posted on services, or in closed **user groups**. Posting content, posting goods and services for sale, user-generated content searching and user groups are all included in the risk profiles due to their role in propagating drugs offences.

Encrypted messaging is often favoured over messaging without automatically enabled encryption, due to its added security.

Users often have to connect with potential dealers through **user connections** before viewing their **user profiles** and associated content. It is also common for suspected dealers to connect with one another, which may provide another way for users to find user profiles offering drugs for sale. These functionalities are also included in the risk profiles.

Other functionalities can be used to perpetrate these offences. **Content tagging** can be an effective way for dealers to consolidate and drive potential buyers to their supply, sometimes with coded labels to avoid detection. Prospective buyers are typically directed towards closed channels of communication with added privacy such as **direct messaging**, where they communicate with suppliers. **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.

Introduction

6H.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the drugs and psychoactive substances 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').

6H.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.

6H.3 Although we have not considered specific evidence about the dark web for the purpose of this chapter, we acknowledge that it continues to play a role in the online drug supply market.[775] [776]

6H.4 There is a limited amount of evidence about how these offences manifest online, so we have placed particular emphasis on one source of evidence from Volteface, an advocacy organisation specialising in drug harm reduction.

[775] RAND, 2022. Commission on Combating Synthetic Opioid Trafficking. [accessed 17 October 2022].

[776] European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Demant & Bakken. 2019. Technology-facilitated drug dealing via social media in the Nordic countries. [accessed 17 October 2022].

Relevant offences

6H.5 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to 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:[777]

- Unlawful supply, or offer to supply, of controlled drugs[778]
- Prohibition of supply etc of articles for administering or preparing controlled drugs[779]
- Inciting any offence under the Misuse of Drugs Act[780]
- Supplying, or offering to supply, a psychoactive substance[781]

6H.6 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6H.7 Examples of how these priority offences may manifest online include posting drugs or psychoactive substances for sale, by posting text, emojis or images on social media services. On social media services, those offering to supply drugs or psychoactive substances may do so both on open user profiles and in closed user groups. This often includes directing users to private messaging services.

6H.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** (Volume 5, Chapter 26).

How drugs and psychoactive substances offences manifest online

6H.9 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.

6H.10 To put the risks of harm into context, a study from Volteface in 2019 found that one in four (24%) young people in the UK aged between 16 and 24 had seen illicit drugs advertised for sale on social media services, and that 18 and 19-year-olds were the most likely to see drugs offered or 'advertised' for sale (29%).[782] Cannabis was the most commonly 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/ecstacy (24%).[783] Compared to over-18s (18-to-24 year olds) and the baseline figure, higher proportions of those aged under 18

[777] 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.

[778] Section 4(3) of the Misuse of Drugs Act 1971.

[779] 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.

[780] Section 19 of the Misuse of Drugs Act 1971.

[781] Section 5 of the Psychoactive Substances Act 2016.

[782] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[783] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

saw cannabis for sale (70%), followed by MDMA/ecstacy (35%) cocaine (28%), LSD/acid (24%), magic mushrooms (14%) and Xanax (27%).[784]

6H.11 The effects of illicit drug supply go far wider than affecting those who consume them and include societal concerns such as the impact on mental health and the burden on the National Health Service.[785] [786] [787] The Department for Public Health, in 2018, estimated that the annual cost of illicit drug misuse in the UK to be around £10.7bn.[788]

6H.12 The Department for Science, Innovation and Technology (DSIT) has estimated the cost of drug supply offences which contain an online element of the 'Illegal drugs' harm. DSIT estimated 5,204 annual cases with a total annual social and economic cost of about £0.2bn. However, it acknowledges that this is likely to be an underestimate due to challenges in identifying and recording the online component.[789]

Risks of harm presented by supply or offer to supply controlled drugs and psychoactive substances

6H.13 The supply of drugs and psychoactive substances is facilitated online by enabling suppliers to market their products and connect with potential buyers. While some online drug dealers opt for postal deliveries[790] to complete the supply of drugs, in-person collections are also favoured. Volteface's research highlighted that some online drug suppliers offer a discount online for buyers who complete the deal in person.[791] With some dealers still offering an offline option (deliveries and collections), the risks associated with traditional street dealing remain, and may arguably increase, due to the wider audience of buyers that suppliers can reach through U2U services.[792] [793]

6H.14 Evidence of the risks of harm arising from the supply of drugs and psychoactive substances online is limited. However, it is likely that this offence leads to a similar experience as for those who use drugs obtained by more traditional means. This could include significant

[784] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[785] 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. Source: NHS Digital, 2021. Statistics on Drug Misuse, England 2020. [accessed 2 October 2023].

[786] In 2022 the Government published statistics relating to the 'Children in Need' census; in the year to 31 March 2022, drug misuse concerns relating to the child and a parent were a factor in 92,250 cases. Children in Need is 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. Source: Children In Need Census, 2022. Characteristics of Children in Need. [accessed 2 October 2023].

[787] Children In Need Census, 2022. Characteristics of Children in Need. [accessed 2 October 2023].

[788] Public Health England, 2018. Alcohol and drug prevention, treatment and recovery: why invest? [accessed 2 October 2023].

[789] DSIT estimated £173m annual cost in 2021/22 prices (updated to £185m in 2022/23 prices, Ofcom price update using GDP deflator). Source: Department for Digital, Culture, Media & Sport, 2022. Online Safety Bill – Impact Assessment (IA) (parliament.uk) [accessed 2 October 2023].

[790] 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].

[791] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[792] 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].

[793] 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. Source: CPS, 2022. County Lines Offending. [accessed 2 October 2023].

health risks, including the development of drug use disorders.[794] Drug use disorders increase morbidity and mortality risks, and individuals may suffer from issues in their personal, family, social, educational, and occupational relationships.[795]

6H.15 In 2022, it was estimated that one in 11 adults (9.2%) aged 16 to 59 in England and Wales had taken an illicit drug in the past year, with use higher among 16-24-year-olds (18.6%).[796] In 2020-2021, cannabis, cocaine and ecstasy were the among the main drugs misused by young people who required substance misuse treatment.[797] In 2021, 4,859 deaths related to drug poisoning were recorded in England and Wales, of which 3,060 were identified as drug misuse.[798] In 2020, it was estimated that the total illicit drugs market in the UK was worth £9.4bn a year, with the total cost of drugs to society more than twice this at £19bn.[799]

6H.16 Another risk presented by the online illicit drug market is the risk associated with counterfeit drugs. Substandard, spurious, falsely labelled and counterfeit (SSFFC) medical products are an increasing threat to consumer health. The World Health Organisation (WHO) recognises this as an urgent health challenge for the next generation.[800] [801]

6H.17 Under-18s may be particularly at risk of harm. Evidence has shown that the use of drugs when an individual's brain is not yet fully developed, for example in those aged under 18, could lead to an increased risk of the onset of depression, psychosis and suicidal tendencies.[802] [803]

6H.18 Moreover, young people can 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*".[804]

[794] World Health Organisation, 2022. Drugs (psychoactive). [accessed 24 October 2022].

[795] World Health Organisation, 2022. Drugs (psychoactive). [accessed 24 October 2022].

[796] Office for National Statistics, 2022. Drug misuse in England and Wales: year ending June 2022. [accessed 7 August 2023].

[797] In addition, a significant proportion of young people reported a problem with alcohol (41%) and nicotine (12%). Source: Office for Health Improvement and Disparities, 2022. Young people's substance misuse treatment statistics 2020 to 2021: report. [accessed 8 May 2023].

[798] Office for National Statistics, 2022. Deaths related to drug poisoning in England and Wales: 2021 registrations. [accessed 8 May 2023].

[799] Home Office (Black, C.), 2020. Review of drugs: summary (accessible version). [accessed 28 July 2023]

[800] World Health Organisation, n.d. Substandard and falsified medical products. [accessed 4 October 2023].

[801] 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 services. Source: UK Health Security Agency (O'Connor, R.), 2018. Alprazolam (Xanax): What are the facts? [accessed 4 October 2023].

[802] Gobbi G, Atkin T, Zytynski T., 2019. Association of Cannabis Use in Adolescence and Risk of Depression, Anxiety, and Suicidality in Young Adulthood, *JAMA Psychiatry.* [accessed 10 May 2023].

[803] Kiburi SK, Molebatsi K, Ntlantsana V, Lynskey MT., 2021. Cannabis use in adolescence and risk of psychosis: Are there factors that moderate this relationship? A systematic review and meta-analysis, *Substance Abuse,* 42(42). [accessed 10 May 2023].

[804] Barnardos, 2023. Child exploitation: a hidden crisis. [accessed 4 October 2023].

Evidence of risk factors on user-to-user services

6H.19 We consider that the risk factors listed below are liable to increase the risks of harm relating to the sale or supply of drugs and psychoactive substances. This is also summarised in the grey box at the start of this chapter.

Risk factors: Service types

6H.20 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 messaging services.

Social media services and video-sharing services

6H.21 There is strong evidence to indicate that the offer to supply drugs and psychoactive substances manifests to a significant extent on social media services. These services can allow users to connect and post information about where to obtain drugs and psychoactive substances, as well as posting items for supply. A study conducted in the UK by Volteface found that 24% of young people (those aged 16-24) reported seeing illicit drugs advertised on social media services.[805]

6H.22 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.[806]

6H.23 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 illicit drug supply. For example, some evidence shows that there are a large number of posts on Instagram that advertise the sale of drugs.[807] 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.[808]

6H.24 Similarly, a study analysing posts on Instagram related to various controlled substances and illicit drugs indicated that of the many posts related to these, there were far fewer posts which explicitly included an offer for supply or an offer to purchase the substances.[809] Still,

[805] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].
[806] Moyle et al. 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 17 October 2022].
[807] To find and collect Instagram posts related to 'study drugs' (Prescription stimulants, also termed Study drugs), researchers conducted hashtag searches with a number of relevant terms; #studydrugs, #nootropics, #cognitiveenhancers, #smartdrugs, and #modafinil. A total of 563 posts including a selection of post comments, were included in the study and analysed using the Content Analysis Method. 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. Source: Petersen, M.A., Petersen, I.L., Poulsen, C., Norgaard, L.S., 2021. #studydrugs-Persuasive posting on Instagram, *International Journal of Drug Policy*, 95. [accessed 17 October 2022].
[808] This study collected and analysed five months of Twitter data (from June-November 2015) filtered for the keyword 'fentanyl' using Amazon Web Services. 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. Source: Mackey, T.K. and Kalyanam, J., 2017. Detection of illicit online sales of fentanyls via Twitter. [accessed 17 October 2022].
[809] In a 2018 study, Li *et al* 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

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*".[810]

6H.25 Moreover, a recent study on image searching for Fentanyl-related precursors[811] found that a large number of image search results were sourced from Pinterest (600+ URLs), followed by Facebook (200-300 URLs), LinkedIn (200+ URLs), Twitter (100+ URLs) and Tumblr (0-100 URLs).[812] From this study we can assume that pictures of drugs are being posted on social media services, and it is possible that some of this content might be related to the supply of these drugs.

6H.26 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[813] 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.[814]

Messaging services

6H.27 Messaging services are commonly used to offer and facilitate the supply of drugs, as they offers 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.[815]

6H.28 Direct messaging, a functionality that is central to messaging services, as well as encrypted messaging, were also found to be risk factors.[816] The latter supports the conclusion that messaging services with encryption are an important service type used in the supply of drugs and psychoactive substances.[817]

more information or requesting to buy the drug. Source: Li, J., XU, Q., Shah, N. and Mackey, T.K., 2018. A Machine Learning Approach for the Detection and Characterization of Illicit Drug Dealers on Instagram: Model Evaluation Study. *Journal of Medical Internet Research.* [accessed 17 October 2022].

[810] Li, J., XU, Q., Shah, N. and Mackey, T.K., 2018. A Machine Learning Approach for the Detection and Characterization of Illicit Drug Dealers on Instagram: Model Evaluation Study. *Journal of Medical Internet Research.* [accessed 17 October 2022].

[811] A precursor is a chemical needed to synthesise a drug.

[812] RAND, 2022. Commission on Combating Synthetic Opioid Trafficking. [accessed 17 October 2022].

[813] See Risk factors: functionalities and recommender systems section for more information. Source: Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[814] See Risk factors: functionalities and recommender systems section for more information. Source: Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[815] Source: 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]; 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].

[816] 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]; Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[817] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022]; Moyle et al. 2019. #Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs, [accessed 17 October 2022]; C4ADS, 2020. Lethal exchange: synthetic drug networks in the digital era. [accessed 17 October 2022].

Risk factors: User base

User base demographics

6H.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.

6H.30 Data suggests that user-base characteristics including **age** and **gender** can lead to increased risks of harm.

6H.31 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"*.[818] Other evidence found that under-18s were as likely or more likely than over-18s to see drugs promoted on specific U2U services.[819]

6H.32 Gender cannot be established as a risk factor; however, there is some research suggesting that more men than women report seeing crack cocaine and heroin advertised for sale online,[820] while more women than men had seen Xanax advertised for sale (23% vs 18%).[821]

Risk factors: Functionalities and recommender systems

User identification

User profiles

6H.33 Volteface research found that user profile features were used to promote drugs for sale.[822] 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.[823]

6H.34 The Volteface study 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 and also found 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.[824]

Anonymous user profiles

6H.35 Research has found that anonymous user profiles are instrumental for users suspected of drug dealing. The profile images are often unidentifiable or anonymous, usually showing a drug-related image. The study also observed that dealers have multiple accounts, usually with similar names on their profile.[825]

[818] Calouri, J., Mooney, B. and Kirk, E., 2022. Running out of credit: Mobile phone tech and the birth of county lines. [accessed 12 May 2023].

[819] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[820] 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. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[821] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[822] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[823] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[824] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 10 October 2023].

[825] *"For example, one account had two profiles where the only difference was the surname: 'Green' and 'Greenn'"*. Source: Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

User networking

User connections

6H.36 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.[826]

6H.37 The Volteface report found that dealers often posted heavily with the aim of getting potential customers to notice them and potentially 'follow' them back.[827]

User groups

6H.38 Setting up user groups has been shown to facilitate relevant offences. Closed user groups on some social media services, where an invitation is required, allow suppliers to promote their services, and provides a contact point for potential buyers.

6H.39 Analysis by the Center for Advanced Defense 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.[828]

6H.40 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.[829]

User communication

Direct messaging

6H.41 Direct messaging is a functionality used to offer to supply drugs. For example, an 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 whom to contact and how.[830]

6H.42 Volfteface found that drug dealers promote drugs on their user profiles or other more open forms of communication using captions to photos or videos that would often instruct individuals that they should "*direct message or dm them*". Researchers concluded that "*communication between potential customers and drug dealers would happen privately via the direct messaging function*".[831]

6H.43 The Moyle *et al*. 2019 report found a specific social media service to be one of the most popular services for purchasing drugs and other psychoactive substances. Messages were

[826] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].
[827] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].
[828] C4ADS, 2020. Lethal exchange: synthetic drug networks in the digital era. [accessed 17 October 2022].
[829] 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].
[830] 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].
[831] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

sent from suppliers to their connections in order to offer illicit drugs for sale. Text within ephemeral messages would describe the illicit drugs and how to obtain them, with emojis occasionally used in place of text. Users could message through the service or continue the conversation elsewhere, as indicated by the supplier.[832]

Encrypted messaging

6H.44 Encrypted messaging is a key component in the supply of drugs and psychoactive substances. 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.

6H.45 For example, Facebook Messenger provides a secure messaging service which dealers and buyers use to arrange deals with known suppliers.[833] But the evidence also shows that it is unusual to see dealers signposting users to contact them on Facebook Messenger. Volteface hypothesises that this may be because this service only uses end-to-end encryption if the user turns it on manually. It is more usual for dealers to direct potential buyers to use private messaging services which automatically provide end-to-end encryption, or to phone or text them privately.[834]

6H.46 With regards to direct messaging, Volteface concluded that "*it was common for dealers to navigate potential customers to alternative encrypted methods of communication*".[835]

6H.47 Analysis from the Center for Advanced Defense 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.[836]

Ephemeral messaging

6H.48 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 offer to supply illicit drugs.[837] 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.

6H.49 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 appear to have some traction in online communities. Here, users contrast the insecurity of text messages and phone calls with the 'safety' of this service, which they assume does not store a database of users' photos, videos and text. Survey and interview

[832] 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 pp.101-110. [accessed 3 June 2023].

[833] 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 pp.101-110. [accessed 3 June 2023].

[834] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[835]Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[836] C4ADS, 2020. Lethal exchange: synthetic drug networks in the digital era. [accessed 17 October 2022].

[837] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

data suggest that these features attract not only customers, but also appeal to some commercial drug suppliers which advertise their products on these services and then require buyers to close the deal on encrypted services.[838]

Reacting and commenting on content

6H.50 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.[839]

Posting content (images, videos, text, emojis)

6H.51 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 very 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.[840] Dealers also caption pictures of drugs with the product names.[841] Dealers were also found to post images of block text rather than pictures of the product to promote their produce and prices.[842]

6H.52 An EMCDDA study which analysed individual posts within social media groups found that posts offering drugs dominated the groups' activities. Fifty per cent of the collected posts were offering drugs for sale.[843]

6H.53 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.[844]

6H.54 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.[845]

6H.55 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. Moyle *et al.* found that dealers would send out several messages a day to

[838] 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].

[839] 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].

[840] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[841] "For example, to indicate the strain of cannabis, the account would describe it as 'Lemon Haze'". Source: Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[842] for example: "200 vals £90 200 lorazepam £120". Source: Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[843] The study was a qualitative study using online ethnography (general drug-related searches were conducted on various social media platforms), and semi-structured interviews (n107) across 5 Nordic countries. The qualitative study was supplemented with data analysis from the national Danish Youth Profile Survey of young people (13-25 years old) to establish the prevalence of online versus offline purchasing of drugs and to compare online and offline buyers. 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].

[844] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[845] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

say what products they had or any special offers. [846] Dealers would also 'prove' the quality and legitimacy of their product by posting videos of themselves using the products.[847]

Transactions and offers

Posting goods or services for sale

6H.56 The 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*".[848] The study also found evidence of suspected drug dealers using functions whereby users can post an item for sale.[849]

6H.57 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".[850]

Content exploring

User-generated content searching

6H.58 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[851] when survey respondents were asked which method was easiest for obtaining contact details for a drug dealer.[852]

Content tagging

6H.59 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.[853]

6H.60 The Commission on Combating Synthetic Opioid Trafficking found that most content promoting Fentanyl on Pinterest were labelled by the author with misleading labels. This was understood to be a method for the authors to circumvent automated content moderation.[854]

6H.61 Hashtags are also used to search for potential sellers.[855]

[846] 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].

[847] 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].

[848] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[849] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[850] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[851] Other options asking a friend, asking a family member, and asking a stranger. Source: Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[852] Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

[853] 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 or Medical Internet Research, 20(4). [accessed 17 October 2022].

[854] RAND, 2022. Commission on Combating Synthetic Opioid Trafficking. [accessed 17 October 2022].

[855] 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, pp.101-110. [accessed 3 June 2023].

Hyperlinking

6H.62 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 suspected accounts post URLs that link to either online pharmacies supplying controlled substances, or to social media services, blogs, user forums and affiliate marketing (i.e. an online advertiser which collects fees for redirecting user traffic to e-commerce platforms), to sell prescription opioids.[856]

6H.63 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.[857]

Recommender systems

Network recommender systems

6H.64 Recommender systems have functions beyond suggesting content that a user is likely to find engaging. U2U 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 (e.g., mutual friend or an old school friend) and are likely to want to connect (or reconnect) with. Studies have found that network recommender systems on services can expose 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.

6H.65 The Volteface research indicates that once a user connects with a dealer who 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 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 as a result of mutual user connections.[858]

6H.66 It is important to acknowledge here that the means of sharing content which 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 and commercial profile

6H.67 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[856] 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 or Medical Internet Research, 20(4). [accessed 17 October 2022].

[857] Petersen, M.A., Petersen, I.L., Poulsen, C., Norgaard, L.S., 2021. #studydrugs-Persuasive posting on Instagram, *International Journal of Drug Policy*, 95. [accessed 17 October 2022].

[858] 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"*. Source: Volteface, 2019. DM for details: Selling drugs in the age of social media. [accessed 17 October 2022].

6I. Firearms and other weapons offences

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

Summary analysis for firearms and other weapons offences: how harm manifests online, and risk factors

This chapter summarises the risks of harm to individuals from several offences linked to the buying and selling of firearms and other weapons.

The risks of harm to individuals from these offences are wide-ranging. Supplying or offering to supply firearms and weapons may result in violent crime, with the most extreme consequence being loss of life.

Service type risk factors:

Our evidence points to **marketplaces and listing services** as the most prominent service type in the facilitation or commission of this set of priority offences. Due to their role in propagating this offence, this service type has been included in the risk profiles.

Functionalities and recommender systems risk factors:

The ability to **post goods or services for sale** enables users to sell, market and purchase firearms and weapons, while **user-generated content searching** can allow them to find firearms and weapons. Due to their role in enabling the sale of firearms and weapons, posting goods or services for sale and user-generated content searching are included in the risk profiles.

By association with other buying and selling offences (the supply of drugs and psychoactive substances), it may be possible to infer that **anonymous user profiles**, and **direct messaging** may be risk factors associated with the sale, hire, purchase and marketing of firearms and weapons, particularly as the latter can be used to progress a transaction. This is also true of the ability to **comment on content** and **tag users,** which can enable direct contact between a prospective buyer and seller.

Introduction

6I.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the firearms and other weapons 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').

6I.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.

6I.3 The firearms and weapons offences cover various matters relating to the online sale etc. of a range of firearms and weapons. In the UK, weapons are classified as restricted or prohibited. Generally speaking, prohibited weapons are subject to the strictest limitations on sale.

6I.4 Although we have not considered specific evidence about the dark web for the purpose of this chapter, we acknowledge that it continues to play a role in the firearms and weapons offences.

Relevant offences

6I.5 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. In regard to firearms and weapons offences, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

6I.6 The priority offences relating to firearms and weapons are the following (note that some of the apparent overlap between these offences is due to similar but separate offences applying in different parts of the United Kingdom):[859]

- purchase etc. of firearms or ammunition without certificate[860]
- purchase etc. of shotgun without certificate
- dealing etc. in firearms or ammunition by way of trade or business without being registered
- sale etc. of firearms or ammunition to person other than registered dealer
- purchase, sale etc. of prohibited weapons
- sale etc. of firearms or ammunition to persons previously convicted of crime
- purchase etc. of firearms or ammunition by person under 18
- supplying firearms to minors
- supplying imitation firearms to minors
- sale and letting on hire of crossbow[861]
- purchase and hiring of crossbow
- sale etc. of offensive weapons[862]
- sale of knives etc. to persons under 18
- sale etc. of knives
- sale of knives etc. to minors
- unlawful marketing of knives[863]
- publication of material in connection with marketing of knives
- sale etc. of firearms or ammunition without certificate[864]
- sale etc. of firearms or ammunition to person without certificate etc
- purchase, sale etc. of prohibited weapons

[859] We are considering offences under section 1(1) or (2) of the Restriction of Offensive Weapons Act 1959.

[860] The offences listed at points (i) to (ix) refer to the following provisions of the Firearms Act 1968: section 1(1); 2(1); 3(1); 3(2); 5(1),(1A) or (2A); 21(5); 22(1), 24; 24(A)

[861] The offences listed at points (x) to (xi) refer to section 1 and 2 respectively of the Crossbows Act 1987

[862] The offences listed at points (xii) to (xiii) refer to section 141(1) or (4); section 141A of the Criminal Justice Act 1988

[863] The offences listed at points (xvi) to (xvii) refer to Section 1 and Section 2 of the Knives Act 1997

[864] The offences listed at points (xviii) to (xxii) refer to Article 24; Article 37(1); Article 45(1) and (2); Article 63)8), Article 66A of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3))

- sale etc. of firearms or ammunition to people who have been in prison etc.
- supplying imitation firearms to minors
- sale etc. of realistic imitation firearms[865]
- requirement for air weapon certificate[866]
- restrictions on sale etc. of air weapons

6I.7 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6I.8 Examples of firearms and weapons offences may include posting them 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.

6I.9 Our evidence shows that U2U services can facilitate the sale and purchase of firearms and weapons by allowing users to post items for sale. 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.[867]

6I.10 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** (Volume 5, Chapter 26).

How firearms and weapons offences manifest online

6I.11 This section is an overview which looks at how firearms and weapons offences manifest online, and how individuals may be a risk of harm.

6I.12 To put the risks of harm into context, according to a National Crime Agency (NCA) assessment, few firearms are sold via the clear web.[868] 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 between the supplier and the buyer.[869]

6I.13 However, investigative reporting from Which?, a consumer protection organisation, highlighted that third-party sellers on large online marketplaces were advertising the sale of various prohibited weapons, including different kinds of knives or other violent tools, to UK consumers.[870]

6I.14 Moreover, law enforcement authorities have discovered a variety of prohibited weapons advertised for sale online and delivery to the UK on a large e-commerce website. Items have included irritant sprays, electric shock devices,[871] front-venting blank-firing firearms and 3D-

[865] An offence under section 36(1)(c) or (d) of the Violent Crime Reduction Act 2006.
[866] The offences listed at points (xxiv) to (xxv) refer to Section 2 and Section 24 of the Air Weapons and Licensing (Scotland) Act 2015 (asp 10).
[867] The Crown Prosecution Service, 2023. Offensive Weapons, Knife Crime Practical Guidance. [accessed 20 September 2023].'
[868] Publicly accessible internet. Often referred to as the 'surface web'.
[869] National Crime Agency, n.d. Illegal firearms. [accessed 20 September 2023].
[870] Which?, 2022. Illegal weapons for sale on AliExpress, Amazon, eBay and Wish, Which? warns. [accessed 20 September 2023].
[871] Commonly referred to as a 'Stun Gun' and by the manufacturer brand 'TASER'.

printed firearms.[872] 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.[873]

6I.15 The ability to share files for the 3D printing of weapons is an emerging way in which U2U services can pose a risk of harm. This evolving threat from 3D-printed firearms[874] has resulted in a recent consultation launched by the Government.[875] The NCA, along with other law enforcement authorities, has highlighted the extent of the threat through recent arrests and convictions. For example, in 2023 three men were convicted for possession of several semi-automatic firearms, printed from files available online.[876] Another example is the use of digital templates to print vital components to convert blank-firing firearms into viable lethal weapons.[877]

Risks of harm to individuals presented by firearms and weapons offences online

6I.16 Supplying or offering to supply firearms and weapons may also result in violent crime, with the most extreme consequence being loss of life. The online sale and subsequent importing of prohibited front-venting blank- firing firearms continues to be a focus for the NCA.[878] This is indicative of the acute risk of harm to individuals from such weapons being marketed and sold online.[879]

6I.17 Indeed, while the NCA's assessment highlights that the level of firearm-related crime in the UK is low, addressing the potential for harm from urban street gangs, organised crime groups and potential terrorists remains a priority for UK law enforcement.[880] [881]

6I.18 In the year ending March 2022, there were approximately 45,000 offences involving a knife.[882] In the same year, 19,555 cautions and convictions were made for possession of a

[872] 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].
[873] Crown Prosecution Service, 2023. Offensive Weapons, Knife Crime Practical Guidance. [accessed 20 September 2023].
[874] GNET, (Basra, R.), 2022. The Future is Now: The Use of 3D-Printed Guns by Extremists and Terrorists. [accessed 20 September 2023].
[875] 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].
[876] West Yorkshire Police, 2023. Update: Third man jailed in 3D printed firearms investigation. [accessed 20 September 2023].
[877] National Crime Agency, 2023. NCA raid suspected gun factory in south London. [accessed 20 September 2023].
[878] Between 2018 and 2022, 550 prohibited weapons and 350 rounds of ammunition were seized in the UK as part of the NCA's focus. Source: Police Professional, 2022 Former fire fighter jailed over gun haul. [accessed 20 September 2023].
[879] National Crime Agency, n.d. Illegal firearms. [accessed 20 September 2023].
[880] 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].
[881] 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].
[882] House of Commons Library, (Allen, G and Burton, M.), 2023. Knife Crime in England and Wales: Statistics [accessed 20 September 2023].

knife or offensive weapon; 18% of the cases involved juveniles aged between 10 and 17.[883] It is important to note that it is not possible to establish how many of these firearms and knives were originally bought or marketed online.

6I.19 The glamourisation of firearms and weapons by 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.[884]

6I.20 Children's exposure and awareness to weapons can be far-reaching, impacting on school absenteeism and physical and mental harm. In a survey by the Youth Endowment Fund 55% of teens said they had seen real-life acts of violence on social media in the past 12 months.[885]

Evidence of risk factors on user-to-user services

6I.21 We consider that the risk factors below are liable to increase the risks of harm relating to firearms and weapons offences. This is also summarised in the grey box at the start of the chapter.

Risk factors: Service types

Marketplaces and listing services

6I.22 Our evidence points to online marketplaces and listing services as the most prominent service type used in the facilitation or commission of relevant offences. Discussion forums and chat rooms have also been identified in the evidence.

6I.23 The NCA's National Strategic Assessment of Serious and Organised Crime report, which includes an analysis of the impact of firearms and other weapons in the UK, states that online forums, auction and online marketplaces are online spaces where the trade of illegal firearms takes place in the UK and in many EU countries.[886]

6I.24 An investigation by Which? in 2022 also found that third-party sellers were listing illegal weapons across a variety of popular online marketplaces.[887]

[883] House of Commons Library, (Allen, G and Burton, M.), 2023. Knife Crime in England and Wales: Statistics [accessed 20 September 2023].

[884] This study was among a small sample of around 20 youths aged 18-24 in Scotland. Source: 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. [accessed 20 September 2023].

[27] Youth Endowment Fund, 2022. Children, violence and vulnerability 2022. [accessed 20 September 2023].

[886] The Dark Web, which is out of scope of the Act, is also named as an online space by the NCA. Source: National Crime Agency, n.d. Illegal firearms. [accessed 20 September 2023].

[887] Which?, 2022. Illegal weapons for sale on AliExpress, Amazon, eBay and Wish, Which? warns. [accessed 20 September 2023].

Risk factors: User base

User base demographics

6I.25 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.

6I.26 Beyond the evident risk of harm to victims that arises from access to and use of firearms and weapons in the UK, there is some indication that children and young people in particular can inadvertently or accidentally be exposed to this harm, as referenced above.[888] This suggests that the **age of users** may be a risk factor.

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles

6I.27 Our research suggests that anonymity is 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.[889] [890] From this it can be inferred that anonymous user profiles, which can allow users to operate anonymously, are a risk factor.

Transactions and offers

Posting goods or services for sale

6I.28 As mentioned above, an investigation by Which? found that firearms 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 starting from as little as 49p*".[891]

Direct messaging

6I.29 From assessing other buying and selling offences, such as chapter 6H: Drugs and psychoactive substances offences, it may be possible to infer that direct messaging may be a risk factor associated with the sale, hire, purchase and marketing of firearms and weapons. Direct messaging can be used to progress a transaction between the buyer and the seller.[892]

[888] Youth Endowment Fund, 2022. Children, violence and vulnerability 2022. [accessed 20 September 2023].

[889] Bakken, S.A and Demant, J.J., 2019. Sellers' risk perceptions in public and private social media drug markets, *International Journal of Drug Policy*, 73. [accessed 20 September 2023].

[890] 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 20 September 2023].

[891] Which?, 2022. Illegal weapons for sale on AliExpress, Amazon, eBay and Wish, Which? warns. [accessed 20 September 2023].

[892] 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 20 September 2023].

Commenting on content and user tagging

6I.30 As detailed in chapter 6H: Drugs and psychoactive substances offences, direct contact between a prospective buyer and seller can be enabled via the ‘comments’ related to a post. This could also apply to the supply of firearms and other weapons. Comments can also enable a further level of connectivity, via the ability to tag a seller to bring a specific comment to their attention.[893]

Content exploring

User generated content searching

6I.31 The same 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.[894]

Risk factors: Business models and commercial profiles

6I.32 No specific evidence was found on how business models may influence risks of harm to individuals for these offences.

[893] Volteface, 2019. DM for Details: Selling Drugs in the Age of Social Media. [accessed 20 September 2023].

[894] Which?, 2022. Illegal weapons for sale on AliExpress, Amazon, eBay and Wish, Which? warns. [accessed 20 September 2023].

6J. Unlawful immigration and human trafficking offences

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

Summary analysis for unlawful immigration and human trafficking offences: how harm manifests online, and risk factors

This chapter covers offences relating to unlawful immigration and human trafficking.

The risks of harm from unlawful immigration can be widespread, including trauma, financial hardship, injury or even death for those undertaking potentially dangerous routes of entry.

An individual's experience of harms from human trafficking offences are unique to their situation, but survivors of these offences can be considered one of the most vulnerable groups at risk of complex mental health difficulties, as well as a number of long-lasting physical health problems.

Service type risk factors:

Social media services and **messaging services** are shown to be risk factors for both the unlawful immigration and human trafficking offences. Among other elements, they are used to target potential victims and advertise services. Often, perpetrators will identify their victim on social media before transitioning to an encrypted private messaging service to further the facilitation of the offence. Due to their role in propagating these offences, social media services and private messaging services are included in the risk profiles.

Adult services are also recognised as risky in the context of the human trafficking offences, sexual exploitation in particular.

Further evidence suggests that services which enable users to **build online communities** can be used by both smugglers and traffickers to target potential victims and share information among themselves.

Functionalities and recommender systems risk factors:

In the case of both unlawful immigration and human trafficking offences, **user profiles** can be exploited by a perpetrator looking to build trust with their victim. Closed **user groups** were found to be a risk factor, with smugglers and traffickers using these groups to share information which could help facilitate the offences. The functionalities of user profiles and user groups have been included in the risk profiles because of their use in perpetrating these offences.

Posting content is also used to promote illegal services in both unlawful immigration and human trafficking offences. Within the unlawful immigration offences, posting content could also be used by a smuggler to build trust in the journey's safety. **Fake profiles** may also be used to manipulate a victim further under human trafficking offences.

Within human trafficking offences, potential traffickers post content which glamourises drug-dealing lifestyles and provides the starting point for online county lines exploitation. The ability to **post goods and services for sale** can allow perpetrators to make fake job offerings to target victims; this is also a recognised tactic within the evidence.

Business model risk factors:

There is very limited evidence on how different revenue models can affect the unlawful immigration and human trafficking offences. Nevertheless, services providing classified listings or other advertising opportunities may increase risk, as the services are incentivised to maximise advertising revenues, and 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.

Introduction

6J.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to unlawful immigration and human trafficking 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').

6J.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.

6J.3 The evidence presented below focuses on unlawful immigration and human trafficking offences on U2U services. Where we refer to 'smuggling', this generally relates to the unlawful immigration offences, and 'trafficking' relates to the human trafficking offences.[895] It is important to recognise that these are harms and offences which are also experienced by children.

6J.4 Ofcom has reviewed the available evidence to develop this assessment of the risks of unlawful immigration and human trafficking offences 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.

[895] '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].

Relevant offences

6J.5 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to unlawful immigration and human trafficking, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act:

- illegal entry and similar offences[896]
- assisting unlawful immigration[897]
- arranging or facilitating the travel of another person, or taking a relevant action, with a view to them being exploited (human trafficking)[898]

6J.6 The Act also covers inchoate offences such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6J.7 The unlawful **immigration offences** covered in this chapter are the 'illegal entry and similar offences' and 'unlawful immigration' listed in (i) and (ii) above. 'Illegal entry' means an individual entering the United Kingdom in breach of a deportation order, entering without permission to remain,[899] 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 fostered 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.

6J.8 The **human trafficking offences** covered in this chapter relate specifically to human trafficking, which includes transporting, recruiting or harbouring an individual with a view to their being exploited. Human trafficking offences 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 (see Chapter 6C: Child sexual exploitation and abuse).

6J.9 Online aspects of human trafficking offences can present as advertisements for job opportunities to recruit people into being trafficked. People are trafficked for numerous reasons, including sexual exploitation. This chapter will include evidence of the role of the internet in trafficking people online, regardless of whether the person is trafficked for slavery or sexual exploitation. (See Chapter 6C: Child sexual exploitation and abuse and Chapter 6K: Sexual exploitation of adults for further details on sexual exploitation.)

6J.10 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** (Volume 5, Chapter 26).

[896] Section 24(A1), (B1), (C1) or (D1) of the Immigration Act 1971.

[897] Section 25 of the Immigration Act 1971.

[898] 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.)).

[899] Generally referred to as 'leave to enter or remain'. Source: Home Office, 2023. Immigration Rules. [accessed 25 September 2023].

How unlawful immigration and human trafficking offences manifest online

6J.11 This section is an overview which looks at how unlawful immigration and human trafficking offences manifest online, and how individuals may be at risk of harm.

6J.12 **Unlawful immigration**: It is not possible to determine the total number of migrants who have come to the UK illegally.[900] Nonetheless, government statistics report an increase over the past few years. In 2022, 45,755 people were reportedly detected arriving. Fifty-one per cent of these people arrived in the three months from August to October, and in August 2022 there was the highest number of arrivals in any month since data has been collected (8,631). The total in 2022 was 60% higher than in 2021.[901] **We are aware that some of these people may since have claimed asylum in the UK.** It should be noted that there are legal routes for entering the UK with the intention of claiming asylum. The research and guidance we 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.

6J.13 **Human trafficking:** There were 16,938 potential victims of modern slavery (including cases of human trafficking) referred to the Home Office in 2022, representing a 33% increase on the preceding year (12,706) and the highest annual number since the National Referral Mechanism (NRM) began in 2009.[902]

6J.14 Save the Children defines child trafficking as follows: "Child trafficking refers to the exploitation of girls and boys, primarily for forced labour and sexual exploitation, children account for 27% of all human trafficking victims worldwide, and two in three child victims are girls".[903] Children are more easily placed in situations of danger by smugglers and forced into participation in criminal activities.[904] They are trafficked into and out of the UK to be exploited, as well as within the UK, from one house to another.

6J.15 There is evidence to suggest that online spaces are increasingly being used by organised criminals to facilitate the travel of migrants. The NCA has worked with social media services to tackle a key driver of these offences, which is online organised immigration crime. A collaboration between the NCA and Twitter, TikTok, Instagram, Facebook and YouTube reported in 2023 that more than 3,300 posts, pages or accounts had been suspended since the NCA's inception in 2021.[905] In August 2023, the NCA announced that this collaboration with social media services would be extended, alongside the development of a new 'Online Capability Centre' increasing the capacity and capability of law enforcement to identify this content online.[906]

[900] 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.

[901] 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 Source: Home Office, 2023. Irregular migration to the UK, year ending December 2022. [accessed 25 September 2023]; Home Office, 2023. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2022. [accessed 25 September 2023].

[902] Home Office, 2023. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of uear summary 2022. [accessed 9 October 2023].

[903] Save the Children, n.d. The Fight Against Child Trafficking. [accessed 25 September 2023].

[904] UNODC, 2019. Children on the move, smuggling and trafficking. [accessed 25 September 2023].

[905] National Crime Agency, 2023. NCA and social media companies work together to tackle organised immigration crime. 24 April. [accessed 25 September 2023].

[906] Home Office, 2023. New tech partnerships to stop the boats. [accessed 25 September 2023].

Risks of harm to individuals presented by unlawful immigration and human trafficking offences

Unlawful immigration

6J.16 An Amber Alert report by the NCA in June 2021 explained that the most common types of illegal content relating to these offences on social media services were*: "advertisements for people-smuggling services, the sale of false documents, and combinations of these two services as…the most common types of OIC (organised immigration crime)-related material on social media*".[907] The report goes on to say there are "*a number of indicators that strongly suggest that a post or account is linked to online immigration crime facilitation*". These 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.[908]

6J.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.[909]

6J.18 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, the risk of individuals being scammed by smugglers,[910] and ultimately the risk of death or serious injury. A 2022 Home Affairs Committee report described how "*at least 166 people have died or gone missing in the English Channel since 2014*".[911]

Human trafficking

6J.19 Human trafficking offences also present multiple risks to both children and adults; the victims can be entrapped, trafficked or exploited in many different ways. [912] [913] Government statistics show that 16,938 potential victims of modern slavery[914] were referred to the Home Office in 2022.[915] 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)*".[916]

[907] NCA, 2021. Amber Alert (Not publicly available – used with permission from the NCA).

[908] NCA, 2021. Amber Alert (Not publicly available – used with permission from the NCA).

[909] 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].

[910] 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].

[911] House of Commons, 2022. Channels, crossings, migration and asylum. [accessed 25 September 2023].

[912] 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].

[913] 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].

[914] This includes human trafficking, slavery, servitude or forced labour, as set out in the Modern Slavery Act 2015.

[915] Home Office, 2023. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2022. [accessed 9 October 2023].

[916] 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].

6J.20 Trafficking can have both short- and long-term impacts for children as well, including damage to their physical and mental health, limited or no access to education and drug and alcohol use.[917]

6J.21 Three of the most common areas of human trafficking and exploitation are explored below.

Labour exploitation

6J.22 The 2022 Government statistics show that overall, potential labour exploitation victims accounted for 30% of all referrals into the NRM.[918] A UN report from 2022 found that trafficking for forced labour was more likely than trafficking for sexual exploitation to be carried out by criminal groups.[919]

6J.23 The Metropolitan Police explains how victims can be forced to work very long hours for little or no pay, and are often kept, and work, in terrible conditions.[920] An example from an NCA investigation in 2023 found three suspected victims 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*".[921]

6J.24 While it is challenging to quantify the social and economic cost of the harm, the UK Government presented an estimate for 'modern slavery' to give an indication of the potential scale. The estimated annual cost of modern slavery per victim is approximately £0.4m based on 2021/2022 prices, which covers the costs of physical and emotional harm, the cost of lost output and time, costs to health services, costs to victim services and law enforcement costs.[922]

Sexual exploitation

6J.25 There is evidence of people exploited into sex work who have been forced or tricked by other individuals, sometimes traffickers. These individuals 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.[923] 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 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.[924]

[917] NSPCC, 2023. Protecting children from trafficking and modern slavery. [accessed 25 September 2023].
[918] Home Office, 2023. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2022. [accessed 25 September 2023].
[919] United Nations Office on Drugs and Crime, 2022. Global Report for Trafficking in Persons. [accessed 25 September 2023].
[920] Metropolitan Police, n.d. Modern Slavery. [accessed 25 September 2023].
[921] NCA, 2023. NCA targets crime group suspected of operating slave labour cannabis farms. 25 January. [accessed 25 September 2023].
[922] They estimated 29 offences of 'Modern Slavery' with an online element in 2020/2021, leading to a total annual social and economic cost of £10.7m; however they emphasised that this is likely to underestimate the true scale and total cost of modern slavery. Source: Department for Digital, Culture, Media & Sport, 2022. The Online Safety Bill. [accessed 25 September 2023].
[923] Human Trafficking Institute, 2021. Federal Human Trafficking Report 2020. [accessed 25 September 2023].
[924] Human Trafficking Institute, 2021. Federal Human Trafficking Report 2020. [accessed 25 September 2023].

6J.26 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.[925]

6J.27 Sexual exploitation of a child is complex, with no one linear manifestation either online or offline. In 2020, of the 2,874 reported referrals to the NRM, the victims were predominantly children trafficked within the UK. Of these referrals, 419 children had primarily been victims of sexual exploitation. 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.[926] Child sexual exploitation does not happen exclusively within child trafficking and smuggling; further manifestations are explored below in the county lines section and in chapter 6C: Child sexual exploitation and abuse (CSEA) offences.

6J.28 Evidence has found that those who have been sexually exploited may be subjected to physical, sexual and psychological violence. This can lead to harm to their physical health, including HIV infections, gynaecological problems, substance and alcohol abuse and long-term physical injury.[927] It also has an impact on their mental health; effects include anxiety, depression, self-harm and post-traumatic stress disorder.[928] Sexual exploitation can also create issues for relationships and care-giving. More information on risks associated with sexual exploitation can be found in chapter 6K: Sexual exploitation of adults offences.

County lines[929] exploitation

6J.29 The evidence also suggests a rapid increase in 'county lines' criminal exploitation. In 2022, 2,281 county lines referrals[930] were flagged, accounting for 13% of all referrals received in the year. The majority of these referrals were for boys aged under 17.[931]

6J.30 The wider impacts 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.[932]

[925] Department for Education, 2017. Child sexual exploitation. [accessed 25 September 2023].

[926] Independent Anti-Slavery Commissioner, 2021. Child trafficking in the UK 2021: a snapshot. [accessed 25 September 2023].

[927] McQuaid, J. 2020 Understanding the psychological effects of sex trafficking to inform service delivery. [accessed 25 September 2023].

[928] McQuaid, J. 2020 Understanding the psychological effects of sex trafficking to inform service delivery. [accessed 25 September 2023].

[929] 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. Source: National Crime Agency, n.d. County Lines. [Accessed 12 May 2023].

[930] 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, 2023. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2022. [accessed 18 October 2023].

[931] Home Office, 2023. Modern Slavery: National Referral Mechanism and Duty to Notify statistics UK, end of year summary 2022. [accessed 25 September 2023].

[932] Public Health England, 2021. County Lines exploitation: applying All Our Health. [accessed 12 May 2023].

6J.31 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 on 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 and/or friends being drawn into county lines to help pay off the debt.[933] 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.[934]

Evidence of risk factors on user-to-user services

6J.32 We consider that the risk factors below are liable to increase the risks of harm relating to the unlawful immigration and human trafficking offences. This is also summarised in the grey box at the start of the chapter.

Risk factors: Service types

6J.33 Research indicates that social media services, private messaging services and online adult services can be used to commit or facilitate offences related to unlawful immigration and human trafficking offences.

Social media services and messaging services

6J.34 Social media services are a risk factor for unlawful immigration offences.[935] In the first five months of 2020, the National Crime Agency reportedly referred over 1,200 web pages related to organised immigration crime to social media services for closure.[936] It was highlighted that social 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"*.[937] In 2020, the NCA reported that "*social media was a major tool used in people-smuggling operations, particularly as people reached the final leg of their journey into the UK"*.[938]

6J.35 In 2021, Europol found a total of 455 social media accounts facilitating unlawful immigration from Belarus to Europe. The use of private messaging services is also highlighted: *"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"*.[939]

[933] Rescue and response, 2019. Rescue and response county lines project. [accessed 12 May 2023].
[934] Public Health England, 2021. County Lines exploitation: applying All Our Health. [accessed 12 May 2023].
[935] 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].
[936] Gentleman, A, 2020. Social media refuse to pull people-smuggling pages, MPs told, *The Guardian*, 3 September. [accessed 25 September 2023].
[937] Gentleman, A, 2020. Social media refuse to pull people-smuggling pages, MPs told, *The Guardian*, 3 September. [accessed 25 September 2023].
[938] From evidence submitted to this report. Source: House of Commons, 2022. Channel crossings, migration and asylum. [accessed 25 September 2023].
[939] Europol, n.d. Europol coordinates referral action targeting migrant smuggling from Belarus. [accessed 25 September 2023].

6J.36 Evidence suggests that smugglers will then encourage their target to move to messaging services, usually with encryption.[940]

6J.37 The evidence also suggests that social media services can be used to commit or facilitate human trafficking offences. Traffickers often 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*".[941] Victims can then be led to other services such as job boards that have fake advertisements, or private messaging services.[942] 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.[943]

6J.38 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.[944] 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 offences.[945]

6J.39 Further, the London Rescue and Response County Lines Project (R&R) identifies that social media is a key facilitator of grooming for, and recruitment into, illicit drug enterprises.[946]

Adult services

6J.40 Adult services were also found to be a risk factor in human trafficking offences. In the NCA's modern slavery report, they found that online adult services may unwittingly play a role in *"expanding offenders' client bases"*.[947]

[940]Diba, P., Papanicolaou, G. & Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK [accessed 25 September 2023].

[941] Administration for children and families (Contreras, J. and Chon, K.), 2022. Technology's complicated relationship with human trafficking. [accessed 25 September 2023].

[942] Administration for children and families (Contreras, J. and Chon, K.), 2022. Technology's complicated relationship with human trafficking. [accessed 25 September 2023].

[943]*"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. Technology's complicated relationship with human trafficking. [accessed 25 September 2023].

[944] 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].

[945] Alliance to Counter Crime Online, n.d. Human trafficking: How Social media Fuels Modern Day Slavery. [accessed 25 September 2023].

[946]Rescue and response, 2019. Rescue and response county lines project. [accessed 12 May 2023].

[947] NCA, n.d. Modern slavery and human trafficking. [accessed 25 September 2023].

6J.42 The Bingham Centre describes how technological developments over the past decade has created new opportunities for sexual exploitation. It states that *"Adult service websites, where most sexual services are advertised, negotiated and facilitated in the UK, have been identified as a space where offenders and traffickers can manipulate, entrap, coerce and force individuals into sexual services".*[948]

Services enabling users to build online communities

6J.43 Service types which foster community building are at risk of enabling both unlawful immigration offences and human trafficking 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.[949] Also, evidence shows that smugglers use private groups to share information about crossing itineraries and departure points.[950] This can include information on routes, border closures, transport services and the cost of arranging trips.[951]

[948] Modern Slavery & Human Rights, n.d. The role of adult service websites in addressing modern slavery. [accessed 25 September 2023].

[949] FATF, 2022. Money Laundering and Terrorist Financing Risks Arising from Migrant Smuggling. [accessed 10 October 2023].

[950] Infomigrants (Alboz, D.), 2016. Social media networks, the best friend of smugglers. [accessed 25 September 2023].

[951] Diba, P., Papanicolaou, G. & Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK [accessed 25 September 2023].

6K. Sexual exploitation of adults offences

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 document, 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.

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.

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 impact 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. Due to their role in propagating this offence, these two types of services are included in the risk profiles.

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. An individual's **immigration status** may also be exploited by a perpetrator.

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’. The functionalities of posting goods and services for sale, encrypted messaging and user profiles are therefore included in the risk profiles.

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.

Introduction

6K.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the sexual exploitation 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’).

6K.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.

6K.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 chapter 6J: Unlawful immigration and human trafficking offences.

Relevant offences

6K.4 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to 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.

6K.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[952]
- Controlling a prostitute for gain[953]

6K.6 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6K.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.

6K.8 'Sexual exploitation of adults' will be used in this chapter as an umbrella term for these offences. It is possible that the evidence provided in this chapter will capture a broader range of issues and harm than the specific elements of the offences contained within the Act, but they have been included to help services identify potential risks.

6K.9 For more details on these offences and how services can assess whether content amounts to illegal content, please refer to the **Illegal Content Judgements Guidance** (Volume 5, Chapter 26).

How sexual exploitation of adults manifests online

6K.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.

6K.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 that 6.3 million people are currently experiencing forced commercial sexual exploitation, and that nearly four out of five are women or girls.[954]

6K.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.[955]

6K.13 In 2018, the Home Office published a report detailing the economic and social costs of modern slavery[956] and outlined the estimated unit costs of sexual exploitation. The total unit cost of modern slavery consists of six components: physical and emotional harm, healthcare services, law enforcement costs, anticipation, lost output and victim services. In total, the

[952] 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)).
[953] 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)).
[954] International Labour Organization (ILO), 2022. Global Estimates of Modern Slavery: Forced Labour and Forced Marriage. [accessed 5 July 2023].
[955] 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].
[956] Home Office, 2018. The Economic and Social Costs of Modern Slavery. [accessed 20 September 2023].

total cost per instance/unit cost of sexual exploitation is estimated to be £319,500 in 2016/17 prices. The total cost of sexual exploitation is calculated by multiplying the unit cost of sexual exploitation by the total number of recorded instances of sexual exploitation, and is estimated to be £1,343m -£1,730m per year in 2016/17 prices.

Risks of harm to individuals presented by offences relating to the sexual exploitation of adults

Causing or inciting prostitution for gain

6K.14 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.

6K.15 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.[957]

6K.16 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.[958]

6K.17 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.

6K.18 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.[959] It also has an impact on their mental health, with effects including anxiety, depression, self-harm and post-traumatic stress disorder.[960] Sexual exploitation can also create negative issues for relationships and caregiving.

Controlling a prostitute for gain

6K.19 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.

[957] Human Trafficking Institute, 2021. Federal Human Trafficking Report 2020. [accessed 5 July 2023].

[958] Human Trafficking Institute, 2021. Federal Human Trafficking Report 2020. [accessed 5 July 2023].

[959] McQuaid, J., 2020. Understanding the psychological effects of sex trafficking to inform service delivery, *Forced Migration Review*. [accessed 5 July 2023].

[960] McQuaid, J., 2020. Understanding the psychological effects of sex trafficking to inform service delivery, *Forced Migration Review*. [accessed 5 July 2023].

6K.20 Although the transition of sex workers to legally advertise online has benefits for sex workers' safety, it also has the effect that there is less public visibility of the sex working community, which can lead to opportunities for organised criminals to profit from their exploitation.[961]

6K.21 For those who identify as a sex worker, sexual exploitation manifests differently. 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 comply.[962]

6K.22 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 (see paragraph 6K.18).

Evidence of risk factors on U2U services

6K.23 We consider that the risk factors below are liable to increase the risks of harm relating to the sexual exploitation of adults. This is also summarised in the grey box at the start of the chapter.

Risk factors: Service types

6K.24 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,[963] social media services, video-sharing services, and messaging services.

Social media services, marketplaces and listings services

6K.25 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.[964] 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.[965]

6K.26 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 chapter 6J: Unlawful immigration and human trafficking offences.

[961] National Police Chief's Council, 2019. National Policing Sex Work and Prostitution Guidance 2019. [accessed 5 July 2023].

[962] Stop the Traffik, 2021. Sex Work and Exploitation: What do you need to know?. [accessed 5 July 2023].

[963] 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.

[964] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons 2020. [accessed 5 July 2023].

[965] 2020 United Nations Office Drugs and Crime report.

6K.27 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.*"[966] This will probably also be the case for other victims and survivors who are being sexually exploited.

Messaging services

6K.28 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.[967] There is additional research to suggest that messaging services with encryption are a risk factor.[968]

Video-sharing services

6K.29 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'.[969]

Risk factors: User base

User base demographics

6K.30 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.

6K.31 Data suggests that user base characteristics including **age, gender, disability, language, immigration status, media literacy** and **socio-economic background** of users could lead to an increased risk of harm to individuals.

6K.32 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 commercial sexual exploitation and the completion of higher education, suggesting that lower academic attainment may also be a risk factor.[970]

6K.33 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

[966] 2020 United Nations Office Drugs and Crime report.
[967] Thorn, 2018. Survivor Insights: The role of technology in domestic minor sex trafficking. [accessed 5 July 2023].
[968] See Risk factors: functionalities and recommender systems section for more information.
[969] See Risk factors: functionalities and recommender systems section for more information.
[970] 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.

exploitation, and that nearly four out of five are women or girls.[971] The National Police Chiefs' Council reported that 91% of sexual exploitation impacts women and girls.[972]

6K.34 Evidence suggests that language barriers could make people more vulnerable to people seeking to sexually exploit them.[973]

6K.35 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.

6K.36 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.[974] Similarly, individual's immigration status can make them more vulnerable to sexual exploitation.[975]

6K.37 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.[976]

Risk factors: Functionalities and recommender systems

User identification

User profiles and fake user profiles

6K.38 User profiles are used by those seeking to sexually exploit others as a recruitment tool. Perpetrators will use information provided in user profiles to identify vulnerable people, as well as creating their own user profiles to create appeal when engaging with them 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 chapter 6J: Unlawful immigration and human trafficking offences.

6K.39 It is possible that potential perpetrators use fake user profiles to falsely present themselves or the person they are exploiting. 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 be used by an individual to portray themselves as offering legitimate services that are in fact sexual exploitation.

[971] ILO, 2022. Global Estimates of Modern Slavery: Forced Labour and Forced Marriage. [accessed 5 July 2023].

[972] National Police Chiefs' Council, 2023. Violence Against Women and Girls: Strategic Threat Risk Assessment. [accessed 5 July 2023].

[973] Preventing Exploitation Toolkit, n.d. Communication difficulties. [accessed 5 July 2023].

[974] Preventing Exploitation Toolkit, n.d. Communication difficulties. [accessed 5 July 2023].

[975] International Organization for Migration (IOM), 2019. Handbook on Protection and assistance to migrants vulnerable to violence, exploitation and abuse. [accessed 5 July 2023].

[976] Preventing Exploitation Toolkit, n.d. Communication difficulties. [accessed 5 July 2023].

User communication

Livestreaming

6K.40 Livestreaming has been found to be 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.[977]

Direct messaging and encrypted messaging

6K.41 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.[978] This use of direct messaging to arrange purchase is a growing trend.

6K.42 The added security and privacy offered by encrypted messaging enables abusers, having made contact with a victim, to entice, manipulate, entrap and exploit them into sexual activities for their own 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.[979] 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.’

6K.43 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 messages with a client allows an adult sex worker to assess safety concerns they have about the client before meeting in person.

Posting or sending location information

6K.44 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 as a way to control others. This is explained in more detail in chapter 6G: Controlling or coercive behaviour offence.

6K.45 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.[980]

[977] 2020 United Nations Office Drugs and Crime report.
[978] Thorn, 2018. Survivor Insights: The role of technology in domestic minor sex trafficking. [accessed 5 July 2023].
[979] 2020 United Nations Office Drugs and Crime report.
[980] Beyond the Gaze, 2018. Safety and Privacy for Online Sex Workers. [accessed 5 July 2023].

Transactions and offers

Posting goods or services for sale

6K.46 The ability to post goods or services for sale plays a key role in the commission or facilitation of the offence of ‘controlling a prostitute for gain’. While the evidence below can make reference 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.

6K.47 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.’ It is likely that these listings are being placed on open channels, with the aim of maximising 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 a period of 60 days using ‘online advertisement’.[981] It is unclear whether this was the result of paid advertising (see Revenue model section below), or a posting that advertised the victim-survivor.

6K.48 Giommoni and Ikwu analysed over 17,000 advertisements for female sex workers that were *“posted on the largest dedicated platform for sex work services in the UK.”* [982] This, along with the author’s description of these advertisements, suggests that this includes posted content which has the effect of advertising a service by offering it for sale.[983]

6K.49 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% of the advertisements presented two indicators and 1.7% advertisements reported three or more indicators of human trafficking.[984] Although this study focused on victims and survivors who had been sexually exploited via trafficking, it is likely that these same indicators will also be relevant in identifying people who are being sexually exploited but who have not necessarily been trafficked.

6K.50 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.

[981] 2020 United Nations Office Drugs and Crime report.

[982] 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.

[983] 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 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.’

[984] Giommoni, L. and Ikwu, R., 2021. Identifying human trafficking indicators in the UK online sex market*, Trends in Organized Crime*. [accessed 19 September 2023]

6K.51 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*."[985] The ability to post goods or services for sale can also be used by exploiters to recruit people to traffic, for example. This is explained in more detail in chapter 6J: Unlawful immigration and human trafficking offences.

Risk factors: Business models and commercial profiles

Revenue models

Advertising-based models

6K.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.

6K.53 This is supported by a United Nations report[986] 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.[987]

6K.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*".[988] We think that a significant proportion of this is likely to have been presented by trafficked sexual workers.

Risk factors: User base

User base size

6K.55 Across our evidence base, services with a large user base are frequently cited as being used to commit or facilitate both the unlawful immigration and human trafficking offences. Integral to this is the popularity of these services in the countries from which migrants are looking to enter the UK.

[985] Crown Prosecution Office, 2019. Prostitution and Exploitation of Prostitution. [accessed 20 September 2023].

[986] 2020 United Nations Office Drugs and Crime report.

[987] "*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*". 2020 United Nations Office Drugs and Crime report.

[988] Thorn, 2018. Survivor Insights: The role of technology in domestic minor sex trafficking, p38 [accessed 5 July 2023].

6K.56 There is also evidence to suggest that both smugglers and traffickers use services with a large user base to target individuals, 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*".[989]

User base demographics

6K.57 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.

6K.58 Data suggests that user base demographics including **socio-economic factors** and **mental** health could lead to an increased risk of harm to individuals.

6K.59 The UK-based charity Unseen[990] does not identify a typical victim of the 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.[991] 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.[992]

Risk factors: Functionalities and recommender systems

User identification

User profiles and fake user profiles

6K.60 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 crossings are then contacted on social media to ask details about which routes to take and how to approach an illegal crossing.[993]

6K.61 As with the CSEA offences presented in chapter 6C: CSEA, user profiles are also used to exploit others by acting as a tool to identify vulnerable people. A United Nations report[994] 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. The targets of this approach are not random but chosen for specific characteristics that are presumably identified on victims' and survivors' user profiles.

[989] Administration for children and families (Contreras, J. and Chon, K.), 2022. Technology's complicated relationship with human trafficking. [accessed 25 September 2023].

[990] Unseen is a UK charity which provides safehouses and support in the community for survivors of trafficking and modern slavery.

[991]Unseen, n.d. Modern Slavery Facts and Figures. [accessed 25 September 2023].

[992]Alliance to Counter Crime Online, n.d. Human trafficking: How Social media Fuels Modern Day Slavery. [accessed 25 September 2023].

[993] Rest of World (Joles, B.), 2022. Inside the risky world of "Migrant TikTok". [accessed 25 September 2023].

[994] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons. [accessed 25 September 2023].

6K.62 Fake user profiles can also be used by traffickers to hide their genuine identities in order to manipulate others. 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.[995]

User communications

Ephemeral messaging

6K.63 Studies suggest that county lines exploitation is facilitated through service applications which offer ephemeral messaging.[996]

Posting content (videos, emojis)

6K.64 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.

6K.65 Posted content can allow potential perpetrators to promote and market their services to attract potential customers and build trust.[997]

6K.66 This content is typically posted on social media services with images of the transport used and the duration of the journey.[998] Research conducted by BIRN (the Balkan Investigative Reporting Network) found a number of posts promoting or 'advertising' prices, transport used and routes proposed, on a large social media service.[999]

6K.67 Videos and images are often posted that share common words, emojis, flags, etc., 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.[1000] 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."* [1001]

User networking

User groups

6K.68 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.

6K.69 An article by InfoMigrants[1002] uncovered closed groups where smugglers share information about crossing itineraries and departure points, and post images of the boats that will be used.

[995] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons. [accessed 25 September 2023].

[996] 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].

[997] 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.

[998] Diba, P., Papanicolaou, G. & Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK [accessed 25 September 2023].

[999] Sinoruka, F, 2022. Rise in TikTok Ads Among Albanians Selling Smuggling Operations to UK, *Balkan Insight*, 8 August. [accessed 25 September 2023].

[1000] Rest of World (Joles, B.), 2022. Inside the risk world of "Migrant TikTok". [accessed 25 September 2023].

[1001] Alliance to Counter Crime Online, n.d. Human trafficking: How Social media Fuels Modern Day Slavery. [accessed 25 September 2023].

[1002] Infomigrants (Alboz, D.), 2016. Social media networks, the best friend of smugglers. [accessed 25 September 2023].

6K.70 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.[1003]

6K.71 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.[1004]

Transactions and offers

Posting goods or services for sale

6K.72 The ability to post goods and services for sale can be used to promote unlawful immigration services. This is particularly prevalent on online adult service services.[1005]

6K.73 A United Nations report indicates that advertisements are used as tools by exploiters to recruit victims. The report 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.[1006]

Risk factors: Business model and commercial profile

Revenue models

Advertising-based model

6K.74 A report from the United Nations Office on Drugs and Crime[1007] states 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.

[1003] Diba, P., Papanicolaou, G. & Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK [accessed 25 September 2023].

[1004] Diba, P., Papanicolaou, G. & Antonopoulos, G.A., 2019. The digital routes of human smuggling? Evidence from the UK [accessed 25 September 2023].

[1005] Modern Slavery & Human Rights, n.d. The role of adult service websites in addressing modern slavery. [accessed 25 September 2023].

[1006] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons. [accessed 25 September 2023].

[1007] United Nations Office on Drugs and Crime, 2020. Global report on trafficking in persons. [accessed 25 September 2023].

6L. Extreme pornography offence

Warning: this chapter contains content that may be upsetting or distressing, including examples of sexual violence.

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, and fall within injurious or non-consensual categories. 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 extreme pornographic material has been possessed and viewed with child sexual abuse material (CSAM). We therefore draw similarities in risks where appropriate.

Service type risk factors:

Adult services that provide user-generated pornography may be at a higher risk of showing, or setting out that they offer, extreme pornographic content. While the analysis below only applies to online adult services which 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 which has been produced by providers of pornographic content. This service type is included in the risk profiles.

User base risk factors:

Insights about demographic risks concern perpetrators rather than victims or survivors. Crime data indicates that the creation/posting of extreme pornographic content is primarily committed by men, which indicates that **gender** 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 U2U services. The ability to **search for user-generated content** on U2U services may help users find extreme pornography content. Due to this role in propagating extreme pornography, posting content and user-generated content searching are included in the risk profiles.

Other functionalities can also be involved in 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 child sexual abuse material (CSAM – chapter 6C: child sexual exploitation and abuse offences) 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 livestreaming, as livestream sessions can be selected and purchased by users.

Business model risk factors:

The revenue models of some online adult services rely on ensuring a supply of new content to maintain and increase their user base. This applies to both **advertising and subscription-based revenue models**, which can create an incentive for 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

6L.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the extreme pornography 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’).

6L.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.

Relevant offences

6L.3 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to extreme pornography, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

6L.4 In this chapter, we consider the following offence:

- possession of extreme pornography[1008]

6L.5 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of this offence.

6L.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.

6L.7 Extreme pornographic content includes realistic 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.

6L.8 For more details on the offence and how services can assess whether content amounts to illegal content, please refer to the **Illegal Content Judgements Guidance** (Volume 5, Chapter 26).

How the extreme pornography offence manifests online

6L.9 This section is an overview which looks at how the extreme pornography offence manifests online, and how individuals may be at risks of harm.

6L.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 charging and recording of this offence between 2015 and 2017. [1009] [1010] The most commonly charged category was that of extreme pornography involving an animal.[1011] This is likely to be because it is easier to identify these illegal images as extreme pornographic material.[1012]

6L.11 Moreover, in 2018-19 there were 28 prosecutions for possession of an extreme pornographic image portraying rape/assault by penetration.[1013]

6L.12 Beyond prosecutions, there is little evidence assessing the scale and nature of extreme pornography online. It is hard to establish the likelihood of a user encountering this illegal content. One study by Vera-Gray *et al.* found that one in eight titles (12%) shown to first-

[1008] In England, Wales and Northern Ireland, this falls under section 63 of the Criminal Justice and Immigration Act 2008.

[1009] McGlynn, C. and Bows., H., 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, *The Journal of Criminal Law*, 83(6).

[1010] 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.

[1011] McGlynn, C. and Bows., H. 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, *The Journal of Criminal Law*, 83(6).

[1012] McGlynn, C. and Bows., H. 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, *The Journal of Criminal Law*, 83(6).

[1013] Crown Prosecution Service, 2019. Violence against women and girls report 2018-19. [accessed 4 September 2023].

time viewers on the landing pages of mainstream pornography sites in the UK described a sexual activity that constituted sexual violence.[1014]

6L.13 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.

Risks of harm to individuals presented by online the extreme pornography offence

6L.14 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. 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.

6L.15 As with CSAM (see chapter 6C: CSEA offences for more information), hyperlinks or plain-text URLs to extreme pornographic content are also a way in which users can share extreme pornography.

6L.16 There is no conclusive evidence establishing the impact of harm from possession of extreme pornographic content. This is partly because research on topics such as the possession of extreme pornographic content can be difficult to conduct. Researchers face legal challenges in studying illegal material.[1015]

6L.17 The available evidence tends to focus on the impact of images falling within the violent, injurious or non-consensual categories. In the development of the original extreme pornography offence aspects of legislation pertaining to these categories of images have been underpinned by the notion that representations of this behaviour could incite those viewing the content to act it out themselves.[1016] [1017]

6L.18 However, there is a lack of evidence surrounding the effects of viewing extreme pornography on adults. Studies have typically tried to establish what impact viewing pornography (which is not necessarily extreme in nature) can have on users in general. This has for the most part has been focused on legal pornography and the impact on children, which will be explored further in our risk assessment of content that is harmful to children.

[1014] 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

[1015] 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.

[1016] 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].

[1017] 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.

6L.19 The evidence directly linking consumption of extreme pornography to violent behaviour has been heavily scrutinised.[1018] Several studies have identified weaknesses in the research base that links 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.[1019] Research on this topic is further complicated by the fact that sexually violent acts can be consensual between adults.

6L.20 There is 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.[1020] [1021] 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 suggest a link between the two offences, but the sample is too small to draw any definitive conclusions.

6L.21 The impact on those involved in producing extreme pornographic content, both consensually or non-consensually, is outside the scope of the extreme pornography offence.

6L.22 However, the production of the content is tied to the consumption of the content online. As mentioned in chapter 6C: CSEA offences, if people watch this material online, the market will supply it, which can exacerbate the harm to those involved in its production.

6L.23 Ofcom recognises the negative impacts that the widespread dissemination of content that was created without an individual's consent may have on that person's physical and mental health, including re-traumatisation, depression and anxiety. The knowledge of such content existing may also have long-lasting effects on an individual's ability to form relationships.

6L.24 There are also scenarios of individuals who have created this content consensually but who may still be harmed in its making. For instance, some actors in the pornography sector have expressed negative physical and mental effects as a result of producing pornographic content, with concern raised over what constitutes 'consent' for an actor when filming scenes with violent elements.[1022] [1023]

Evidence of risk factors on user-to-user services

6L.25 We consider that the risk factors below are liable to increase the risks of harm relating to the extreme pornography offence. This is also summarised in the grey box at the start of the chapter.

[1018] McGlynn, C. and Rackley, E., 2009. Criminalising extreme pornography: a lost opportunity, *Criminal law review*, 4, pp. 245-260.

[1019] 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.

[1020] 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).

[1021] 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]

[1022] Cole, S., 2020. A new wave of reckoning is sweeping the porn industry. *VICE*, 10 June. [accessed 4 September 2023].

[1023] 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.

Risk factors: Service types

6L.26 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 adult services can be used to commit or facilitate the extreme pornography offence.

Adult services

6L.27 There is evidence to show that adult 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 pornography sites in the UK described a sexual activity that constituted sexual violence, as defined by the study.[1024] The British Board of Film Classification (BBFC)[1025] 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*".[1026]

6L.28 The study of the landing pages of three online UK adult 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.[1027] 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 adult services, it is possible that these findings could be relevant to other U2U services that allow pornographic content on their services.

Messaging Services

6L.29 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.[1028] There have also been instances of users sending unsolicited extreme pornographic content through messaging apps.[1029]

[1024] 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.

[1025] 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.

[1026] BBFC, 2020. Young people, pornography and age-verification. [accessed 6 September 2023].

[1027] Vera-Gray, F., and McGlynn, C., 2021. Sexually violent pornography is being promoted to first-time users of top sites. [accessed 6 September 2023].

[1028] 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].

[1029] 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].

Risk factors: User base

User base demographics

6L.30 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.

6L.31 Data suggest that it is possible that gender is 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 majority of those charged were men (97%). It is therefore reasonable to assume that more men than women may commit offences relating to extreme pornography.[1030]

Risk factors: Functionalities and recommender systems

User identification

Anonymous user profiles

6L.32 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

6L.33 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 is also of appeal to users seeking CSAM material, it may be possible that user groups are used in similar ways to CSAM (see chapter 6C: CSEA offences for further information). 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

6L.34 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. [1031] Perpetrators may be choosing to use encrypted messaging to avoid detection.

[1030] McGlynn, C. and Bows., H. 2019. Possessing Extreme Pornography: Policing, Prosecutions and the Need for Reform, *The Journal of Criminal Law*, 83(6).

[1031] 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].

Livestreaming

6L.35 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 chapter 6C: CSEA offences and chapter 6B: Terrorism offences, 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)

6L.36 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 adult service, removed 10 million videos, amounting to about 80% of its content, after high-profile coverage raising concerns about the availability of illegal material, including CSAM and non-consensually shared intimate images, hosted on the service.[1032] Christian Action, Research and Education (CARE)[1033] 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*."[1034] 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.[1035]

Transactions and offers

Post goods or services for sale

6L.37 Functionalities that allow users to post goods or services for sale could be used to commit or facilitate extreme pornography offence, 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 chapter 6C: CSEA offences).

Content exploring

User-generated content searching

6L.38 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 adult services found that pornography which contravened the service's own terms and conditions could be found through simple keyword searches.[1036] This suggests that the search functions on online adult 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.

[1032] Concerns were not necessarily raised about extreme pornography material, but primarily child sexual exploitation and abuse material and intimate image abuse (IIA).

[1033] CARE is a Christian public policy charity based in the United Kingdom.

[1034] CARE response to 2022 Ofcom Call for Evidence: First phase of online safety regulation.

[1035] 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.

[1036] Vera-Gray, F. and McGlynn, C., 2021. Sexual Violence in Mainstream Online Pornography. [accessed 6 September 2023].

Building lists or directories

6L.39 As with in-service search functionalities, many online adult 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. It is likely that this curation and categorisation of images will increase the discoverability of such material.

Hyperlinking

6L.40 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. It is possible that 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 pornography and provide a relatively easy user journey to sites showing more niche or even extreme pornographic content.

6L.41 However, the use of hyperlinks to access illegal content is not limited to online adult services. As evidenced in chapter 6C: CSEA (CSAM) offences, 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

6L.42 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

6L.43 The creation of deepfake material[1037] 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 by editing other images. The Illegal Content Judgements Guidance (Volume 5, Chapter 26) has more information on when an altered image should be considered illegal content.

Recommender systems

Content recommender systems

6L.44 Evidence suggests that the design of recommender systems could be a risk factor in the facilitation and commission of the 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 in

[1037] 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.

the Risk factors: Business model section below. 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.[1038]

6L.45 A BBFC study of how young people discover and engage with pornography online found that several boys had found themselves getting lost in a 'rabbit hole' of pornography, where they found increasingly violent videos. [1039] 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.

6L.46 Moreover, as detailed in Risk factors: Service Types, landing pages of online adult 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

6L.47 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, 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 in order to maintain a content offering that will continue to attract users, and that this could result in instances of extreme pornography being presented. [1040]

6L.48 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 make a purchase on a pornographic website with such content.[1041]

[1038] McGlynn, C., and Woods, L., 2022. Image-Based Sexual Abuse, Pornography Platforms and the Digital Services Act. [access 6 September 2023].

[1039] BBFC, 2020. Young people, pornography and age-verification. [accessed 21 September 2023].

[1040] Centre to End All Sexual Exploitation, 2021. Expose Big Porn. [accessed 21 September 2023].

[1041] 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].

6M. 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, of intimate images.

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.

Service type risk factors:

Research indicates that intimate image abuse occurs particularly on **adult services, social media services, and file-storage and file-sharing services.** Due to their role in propagating this offence, these types of services are included in the risk profiles.

Dating services, messaging services, and video-sharing services are also used in perpetration of this offence.

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 18-24. Other user base characteristics have also been found to increase the risk of a user experiencing intimate image abuse: **race and ethnicity, cultural and linguistic diversity, disability and sexual orientation**. Due to its significance to intimate image abuse, and other offences affecting women in particular, **user base demographics** are included as a general risk factor in the risk profiles.

Functionalities and recommender systems risk factors:

Intimate image abuse is primarily committed by **posting images and videos** where perpetrators share intimate images non-consensually. As a result, this functionality is included in the risk profiles.

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 **download content**, such as intimate images, which can then be shared on other services. Due to their role in propagating intimate image abuse, these functionalities are also included in the risk profiles.

Several additional functionalities on 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.

Introduction

6M.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the intimate image abuse 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’).

6M.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.

6M.3 Throughout our risk assessment, we will use the term ‘intimate image abuse’ to refer to the offences covered in this chapter.[1042] This chapter will cover adult intimate abuse; intimate image abuse relating to under-18s is covered in chapter 6C: Child sexual exploitation and abuse offences.

Relevant offences

6M.4 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to intimate image abuse, Ofcom is required to consider the risks of harm connected with the priority offences listed in Schedule 7 of the Act.

[1042] 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’ chapter. These offences are covered separately in chapter 6S: Cyberflashing offence and chapter 6C: Child sexual exploitation and abuse offences.

6M.5 The existing priority offences for intimate image abuse are the following:

- Disclosing, or threatening to disclose, private sexual photographs and films with intent to cause distress – section 33 of the Criminal Justice and Courts Act 2015[1043]
- In Scotland, disclosing, or threatening to disclose an intimate photograph or film – Section 2 of the Abusive Behaviour and Sexual Harm (Scotland) Act 2016

6M.6 There is also a new offence in section 66(B) of the Sexual Offences Act, which (at the time of our consultation) has not yet been brought into force. This consists of a new 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. We consider that the evidence we present in this chapter is largely applicable across the existing and new offences, so will remain relevant if the new offence is brought into force.

6M.7 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6M.8 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 in, participating in, or being present during a sexual act, and/or where their genitals, buttocks or breasts are exposed or covered only with underwear.[1044] 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. For the most part, 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.[1045] Intimate image abuse can form part of a wider continuum of online and offline behaviours by a partner or former partner. These are closely linked to other behaviours explored in greater detail in chapter 6E: Harassment, stalking, threats and abuse offences and chapter 6G: controlling or coercive behaviour offence.

6M.9 Many intimate images are taken or made consensually. However, some intimate image abuse can involve images taken or made without consent. This includes images produced through 'upskirting'[1046] and 'downblousing'[1047], intimate images made as deepfakes[1048], or images taken as a result of hacking or hidden cameras. It also includes intimate images

[1043] Or, if section 188 of the Act is brought into force and Schedule 7 to the Act is amended accordingly before we issue our final document, section 66B of the Sexual Offences Act 2003.

[1044] We are aware that interpretations of what is 'intimate' can vary among different cultural and religious groups, but for present purposes we refer to the new offence in section 66(B) of the Sexual Offences Act 2003. The definition of intimate state 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.

[1045] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 3 August 2023].

[1046] 'Upskirting' refers to someone taking a photograph that appears to have been taken up a person's clothing (such as a skirt) without consent.

[1047] '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].

[1048] 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.

produced through the use of 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 which were made or taken non-consensually, they are within scope of this chapter.

6M.10 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** (Volume 5, Chapter 26).

How intimate image abuse offences manifest online

6M.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 both the non-consensual sharing of intimate images and threats to share intimate images.

6M.12 The intimate images can be shared both consensually and non-consensually. Non-consensual sharing can manifest in the following ways:

- as part of domestic abuse or coercive and controlling behaviour;
- for financial gain, including 'sextortion'[1049];
- and by unknown perpetrators, where the common motivation is a mixture of social status, humour, sexual gratification and misogyny.

6M.13 The latter may include sharing intimate images taken or made through upskirting, downblousing, the creation of deepfakes, screen capture technology, and hidden or hacked cameras. In addition, some preparators will hack into victims' and survivors' accounts to access intimate images.[1050] This is particularly likely when intimate images are shared as part of domestic abuse by an ex-partner, or when they are shared by an anonymous perpetrator. The latter can be seen in high-profile cases of women in the public eye who have been targeted online: their online accounts have been hacked, intimate images of them identified and then shared widely.[1051]

6M.14 Intimate image abuse can be exacerbated by 'collector culture', where groups of users anonymously exchange and swap intimate images of women without their consent.1052 Some groups categorise victims and survivors, predominantly women, by their location.1053

6M.15 Between 2015 and 2021 over 28,000 reports of the disclosure of sexual images without consent were recorded by police.1054 Refuge found that the number of intimate image abuse offences recorded increased by 40% between 2020 and 2021.[1055]

[1049] 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].

[1050] Refuge found that among women who had experienced threats to share intimate images, 10% had been hacked for images. Source: Refuge, 2020. The Naked Threat. [accessed 3 August 2023].

[1051] BBC News, 2016. Celebgate hack: Man to plead guilty to nude photos hack, *BBC News*, 15 March. [accessed 3 August 2023].

[1052] 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].

[1053] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 3 August 2023].

[1054] Ministry of Justice, 2022. New laws to better protect victims from abuse of intimate images. [accessed 3 August 2023].

[1055] Refuge (Bottomely, B. and Bruckmayer, M.), 2023. Intimate image abuse – despite increased reports to the police, charging rates remained low. [accessed 31 August 2023].

6M.16 Intimate image abuse also covers threats to share intimate images. Threats to share intimate images predominantly fall into two categories:[1056]

- as part of a wider pattern of domestic abuse or coercive and controlling behaviour; or
- for financial or other gain, including sharing further intimate images.

6M.17 Threats to share intimate images as part of domestic abuse occur when current or former partners threaten to share intimate images to coerce or exert control over their partner or former partner. This behaviour is often part of a pattern of wider abusive behaviour both offline and online.[1057] See also chapter 6E: Harassment, stalking, threats and abuse offences and chapter 6G: Controlling or coercive behaviour offence to understand the abuse dynamics between partners or former partners that underpin intimate image abuse.

6M.18 In some cases, threats to share are used to blackmail victims and survivors into performing sexual acts or sending more intimate images.[1058] In other cases, threats are made to blackmail victims and survivors into sending money.[1059]

6M.19 There are also financial motivations for intimate image abuse. For example, threats to share intimate images can also often form part of criminal exploitation, through sextortion,[1060] or by gangs threatening to share intimate images to force individuals to take part in criminal activity.[1061]

6M.20 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.[1062] There are indications that instances of threats to share intimate image abuse are rising in the UK. Reports to the Revenge Porn Helpline[1063], which supports all adult victims of intimate image abuse living in the UK, rose from 521 in 2015 to 3,146 in 2020.[1064] Reports of sextortion are rising particularly quickly, as explored further below.[1065]

[1056] Although the majority of threats to share pertain to these two main groups, 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 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].

[1057] 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. The Naked Threat. [accessed 3 August 2023].

[1058] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

[1059] 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 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].

[1060] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023]; Revenge Porn Helpline (Ward, Z.), 2022. Revenge Porn Helpline Report. [accessed 3 August 2023].

[1061] Storrod, M. and Densley, J. 2017. 'Going viral' and 'Going country': the expressive and instrumental activities of street gangs on social media, *Journal of Youth Studies,* 20(6), pp.677-696. [accessed 20 September 2023]; 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), pp.313-329. [accessed 20 September 2023].

[1062] Ministry of Justice, 2022. New laws to better protect victims from abuse of intimate images. [accessed 4 August 2023].

[1063] Revenge Porn Helpline is a UK-based organisation which supports adult victims of intimate image abuse.

[1064] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].

[1065] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].

Risks of harm to individuals presented by intimate image abuse

6M.21 Threats to share intimate images can go on for a significant amount of time, and victims and survivors will often receive multiple threats throughout a single day. This is particularly common where threats form part of a wider pattern of domestic abuse, which may go on for months and years. Regular threats are also often experienced by victims and survivors of threats for financial or other gain.[1066]

6M.22 Threats to share intimate images can have a negative impact on victims and survivors. In particular, threats to share images can cause mental health issues, with considerable distress and anxiety experienced by victims and survivors.[1067] Refuge[1068] found that 83% of women who had been threatened with intimate image abuse said that threats to share intimate images had impacted their mental health and emotional wellbeing.[1069] In some cases, this can lead victims and survivors to consider taking their own life; more than one in ten women who had been threatened with intimate image abuse reported that they had felt suicidal as a result of threats to share intimate images.[1070]

6M.23 Such threats can also affect other areas of victims and survivors' lives. A study by Thorn[1071] looking at the experiences of young people who had been targets of threats to expose sexual images, found that 41% of respondents had lost a relationship with a friend or family member or partner because of the incident. Twelve per cent moved home, 10% of respondents reported problems at school and 8% reported problems in their jobs.[1072] 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.[1073]

6M.24 Research has found that men are the primary perpetrators of intimate image abuse. A study by Cyber Civil Rights Initiative[1074] found that men were twice as likely as women to report being the perpetrators of intimate image abuse.[1075] Where the gender of the perpetrator was known, twice as many men than women were the perpetrators of threats.[1076]

6M.25 Victims and survivors of intimate image abuse describe their experiences as a form of sexual violence – a 'social rupture'. An experience that radically alters their life experiences,

[1066] 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 18-25 year-olds). Source: Thorn (Wolak, J. and Finkelhor, D.), 2016. Sextortion: Findings from a survey of 1,631 victims. [accessed 4 August 2023].

[1067] Robinson, A., 2021. Threats to Share Intimate Images to Become a Crime, *UK Safer Internet Centre,* 6 May. [accessed 4 August 2023].

[1068] Refuge is a UK-based organisation, providing support for women and children experiencing domestic violence.

[1069] Refuge, 2020. The Naked Threat. [accessed 4 August 2023].

[1070] Refuge, 2020. The Naked Threat. [accessed 4 August 2023].

[1071] Thorn is an international non-profit organisation working to develop new technologies to combat online child sexual abuse.

[1072] 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 4 August 2023].

[1073] Refuge, 2020. The Naked Threat. [accessed 4 August 2023].

[1074] 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.

[1075] 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].

[1076] Refuge, 2020. The Naked Threat. [accessed 4 August 2023].

relationship and activities, to such a degree that their lives become divided into 'before' and 'after' the abuse.[1077] Victims and survivors may struggle with a range of feelings on the disclosure of intimate images, including shame, helplessness, self-blame, isolation and humiliation.[1078] 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.[1079]

6M.26 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.[1080] These risks can align with different understandings of what 'intimate' might mean.

Evidence of risk factors on user-to-user services

6M.27 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 in the grey box at the start of the chapter.

Risk factors: Service type

6M.28 Research indicates that the following types of services are used to facilitate or commit offences related to intimate image abuse: adult services, social media services, messaging services, file-storage and file-sharing services, dating services, and video-sharing services.

6M.29 An analysis of 77 services hosting intimate image abuse found content likely to be intimate image abuse across six types of services: dedicated intimate-image abuse sites, user-generated pornography sites[1081], image boards, community forums, blogging platforms and social media services.[1082]

Adult services

6M.30 Adult 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.[1083] A study of mainstream pornography sites found that, of the analysable homepage videos, 2.2% had titles which constituted descriptions of intimate image abuse.[1084] Another study found that an adult site

[1077] 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].

[1078] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

[1079] 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].

[1080] Refuge, 2022. Marked as Unsafe: How online platforms are failing domestic abuse survivors. [accessed 22 August 2023].

[1081] While the analysis only applies to online adult 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.

[1082] 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].

[1083] Revenge Porn Helpline's response to 2021 Law Commission consultation. [accessed 3 August 2023].

[1084] 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].

hosted nearly 60,000 videos under four phrases associated with intimate image abuse.[1085] Additionally, of the intimate images that were reported by the Revenge Porn Hotline in 2020, nearly 50% of the content was found on pornographic sites.[1086]

6M.31 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.[1087]

Social media services

6M.32 There is also evidence of a significant amount of intimate image abuse taking place on social media services. The Revenge Porn Helpline found that Facebook and Instagram accounted for 18% and 15% respectively of reported cases of intimate image abuse in 2020.[1088] 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.[1089] In some cases, intimate images are shared on social media services, and then re-shared to other sites.[1090]

Messaging services

6M.33 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 Hotline highlighted private messages as a source of reported cases of intimate image abuse, representing 18% of cases where images were shared in 2020. This includes private messaging services as well as emails and texts.[1091] A Thorn survey 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.[1092]

File-storage and file-sharing services

6M.34 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.[1093] 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.[1094] The Revenge Porn Helpline

[1085] 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].

[1086] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].

[1087] Office of the eSafety Commissioner, 2017. Image-based abuse. National survey: summary report. [accessed 22 August 2023].

[1088] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].

[1089] Refuge, 2022. Marked as Unsafe: How online platforms are failing domestic abuse survivors. [accessed 22 August 2023].

[1090] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

[1091] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].

[1092] 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].

[1093] McLaughlin, E., 2018. Dropbox removed folder containing explicit photos of female service members, *abc News,* 12 March. [accessed 22 August 2023].

[1094] BBC News, 2019. Victim's warning after finding revenge porn from 'every UK city', *BBC News,* 17 May. [accessed 22 August 2023].

found that file sharing is often part of ‘collector culture’, with downloadable files shared between perpetrators.[1095]

Dating services

6M.35 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.[1096] 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.[1097]

Video-sharing services

6M.36 There is some evidence to suggest that video-sharing services, particularly those with livestreaming, are used to commit intimate image abuse.[1098] Thorn found that for 23% of respondents, perpetrators contacted victims and survivors through video-sharing services.[1099]

6.18

Risk factors: User base

User base size

6M.37 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.

6M.38 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 media services are considered by perpetrators as a place where their material will be seen by people whom the victims and survivors know. The key motivation here for the perpetrator is shaming their target.[1100] 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.

6M.39 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

[1095] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].

[1096] Revenge Porn Helpline, n.d. What to do if you’ve been victim to online webcam blackmail, also known as sextortion. [accessed 12 September].

[1097] 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].

[1098] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

[1099] 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 4 August 2023].

[1100] Office of the eSafety Commissioner, 2017. Image-based abuse. National survey: summary report. [accessed 22 August 2023].

intimate images, particularly those based in countries with less stringent copyright laws. This may be because images are less likely to be taken down.[1101]

User base demographics

6M.40 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.

6M.41 Data suggests that user base characteristics including **gender, age, race and ethnicity, cultural and linguistic diversity, disability, and sexual orientation** could lead to an increased risks of harm to individuals.

6M.42 Intimate image abuse is a gendered harm. Nearly all research on intimate image abuse finds that women are 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 41.9 images per victim. For men the average number of images reported was 1.5 images per victim.[1102]

6M.43 Similarly, the Law Commission identified that victims and survivors of intimate image abuse are primarily women and the perpetrators are primarily men.[1103] As outlined in 6M.21, one type of intimate abuse is strongly tied to domestic violence, with perpetrators trying to control the victim and survivor. Of the 376 prosecutions for intimate image abuse offences commenced in the year ending March 2019, 83% were marked as being domestic abuse-related.[1104]

6M.44 For the most part, threats to share intimate images are disproportionately experienced by women. Refuge estimates that one in seven women aged 18-34 in England and Wales have experienced threats to share their intimate images, in comparison to one in nine men aged 18-34.[1105] 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.[1106] The exception to this is sextortion for money, which is disproportionately experienced by men;[1107] the Revenge Porn helpline estimates that men are five times more likely than women to be a victim of sextortion for money.[1108] 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.[1109]

[1101] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[1102] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].
[1103] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[1104] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[1105] Refuge, 2020. The Naked Threat. [accessed 4 August 2023].
[1106] Cyber Civil Rights Initiative (Eaton, A., Jacobs, H. and Ruvalcaba, Y.), 2017. 2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration: A Summary Report. [accessed 4 August 2023].
[1107] 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].
[1108] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].
[1109] Eaton, A. A., Ramjee, D. and Saunders, J. F., 2022. The Relationship 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].

6M.45 In contrast, threats to share intimate images in order to coerce an individual into sharing more intimate images is predominantly experienced by women.[1110]

6M.46 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, 3% of internet users aged 18-24 had seen or experienced 'sharing of, or threats to share, intimate images without consent' in the past four weeks compared to 1% of all adult internet users.[1111]

6M.47 As with many online harms, the risk of intimate image abuse is likely to be higher for people with protected characteristics.

6M.48 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% vs 1%).[1112] 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.[1113]

6M.49 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.[1114] Similarly, Ofcom research found that more gay or lesbian 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 (2% vs 1%).[1115]

6M.50 Disability can also be a risk factor. Ofcom research found that internet users aged 13+ with any limiting and impacting conditions had experienced the 'sharing of, or threats to share, intimate images without consent' online in the past four weeks more than users with no limiting or impacting conditions (2% vs 1%).[1116]

[1110] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].

[1111] Note: Between waves 1 and 2 the code changed from 'sharing intimate images without consent' to 'sharing of, or threats to share, intimate images without consent'. Source: Ofcom, 2023. Online Experiences Tracker: Data Tables (Waves 1 and 2). [accessed 4 September 2023].

[1112] Note: Between waves 1 and 2 the code changed from 'sharing intimate images without consent' to 'sharing of, or threats to share, intimate images without consent'. Source: Ofcom, 2023. Online Experiences Tracker: Data Tables (Waves 1 and 2). [accessed 4 September 2023].

[1113] 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 Relationship 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].

[1114] 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. Eaton, A., Ramjee, D. and Saunders, J., 2022. The Relationship 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].

[1115] Note: Between waves 1 and 2 the code changed from 'sharing intimate images without consent' to 'sharing of, or threats to share, intimate images without consent'. Ofcom, 2023. Online Experiences Tracker: Data Tables (Waves 1 and 2). [accessed 4 September].

[1116] Note: Between waves 1 and 2 the code changed from 'sharing intimate images without consent' to 'sharing of, or threats to share, intimate images without consent'. Ofcom, 2023. Online Experiences Tracker: Data Tables (Waves 1 and 2). [accessed 4 September].

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

6M.51 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[1117] 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.[1118]

6M.52 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.[1119]

Anonymous user profiles

6M.53 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.[1120]

6M.54 Perpetrators who share the intimate images for the purposes of sexual gratification or financial gain may abuse the anonymity that some services with smaller user bases offer; this makes it less likely that the victims and survivors will realise that their images have been distributed and potentially identify the anonymous perpetrator.[1121]

User networking

User groups

6M.55 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 chapter 6C: CSEA, 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.

[1117] 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.

[1118] eSafety Commissioner, 2019. eSafety for Women from Culturally and Linguistically Diverse Backgrounds: Summary Report. [accessed 22 August 2023].

[1119] Chapter 6E: Harassment, stalking, threats and abuse 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].

[1120] 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].

[1121] Office of the 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 communications

Livestreaming

6M.56 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.[1122]

Direct messaging

6M.57 Direct messaging is a key 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.[1123]

6M.58 It is also likely that direct messaging is used by perpetrators to send online threats to disclose intimate images. As detailed in chapter 6E: Harassment, stalking, threats and abuse, 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.[1124] Other studies suggest that threatening behaviour can result in incessant messaging from perpetrators.[1125]

Encrypted messaging

6M.59 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 were often sent through messaging services which were end-to-end encrypted and not proactively searched by a moderation team.[1126]

Group messaging

6M.60 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.[1127] 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.[1128]

[1122] 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].

[1123] Revenge Porn Helpline (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 4 August 2023].

[1124] 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].

[1125] Refuge, 2021. Unsocial Spaces: make online spaces safer for women and girls. [accessed 22 August 2023].

[1126] Revenge Porn Helpline (Ward, Z.), 2022. Revenge Porn Helpline Report. [accessed 22 August 2023].

[1127] BBC World Service Disinformation Team, 2022. Why won't Telegram take down my naked photos?, *BBC News,* 20 February. [accessed 22 August 2023].

[1128] eSafety Commissioner, 2019. Understanding the attitudes and motivations of adults who engage in image-based abuse. [accessed 22 August 2023].

Posting content (images, videos) and re-posting and forwarding content

6M.61 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.

6M.62 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.

6M.63 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.

6M.64 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.

6M.65 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 adult services.

6M.66 This secondary content sharing can be done using a service's own forwarding functionalities or re-posting functionalities. It can also be done by the user downloading the image to their own device (see Downloading content below) and then sharing it on another service.[1129] As well as re-sharing intimate images to services, it is possible that the intimate images are also being forwarded on private messaging channels.

Transactions and offers

6M.67 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

6M.68 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.

[1129] 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 adult 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].

6M.69 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.[1130] 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.[1131]

6M.70 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 adult 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.[1132] Campaign group #NotYourPorn[1133] has also found an increase in the number of victims and survivors whose intimate images have been tagged or labelled as 'leaked' or 'stolen'.[1134]

Content storage and capture

Downloading content

6M.71 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 'User Communications' above, 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 download content facilitates intimate image abuse as it allows users to possess intimate images which they will then share further.

Screen capturing and recording

6M.72 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.[1135] 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.[1136]

[1130] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[1131] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[1132] 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].
[1133] #NotYourPorn is a UK-based movement focused on protecting non-consenting adults, sex workers and under-18s from image-based sexual abuse.
[1134] Dawson, B., 2020. Revenge Porn Is Being Posted Under a Different Name, *Vice,* 15 December. [accessed 22 August 2023].
[1135] Law Commission, 2021. Intimate Image Abuse: A consultation paper. [accessed 3 August 2023].
[1136] Thorn (Wolak, J. and Finkelhor, D.), 2016. Sextortion: Findings from a survey of 1,631 victims. [accessed 4 August 2023].

Content editing

Editing visual media

6M.73 Functionalities that allow images to be edited may be used to make intimate images without consent. This could include images such as deepfakes[1137] which can subsequently be shared with other users.

Recommender systems

6M.74 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 service 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 the images. The likelihood of posted intimate images being disseminated may increase if they receive sufficient 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

6M.75 There is no known evidence specific to business models and the facilitation or commission of these offences.

[1137] 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.

6N. Proceeds of Crime offences

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 could 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. Further information on fraudulent activity can also be found in chapter 6O: 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.

This chapter primarily focuses on money mule activity, as the evidence base on the manifestation of the offence is more extensive.

The risks of harm to individuals that could arise from proceeds of crime offences include losing livelihoods, losing access to financial services and severe mental health concerns.*Service type risk factors:*

Our evidence shows that proceeds of crime offences committed or facilitated online rely on **social media services** and **messaging services** that have a broad user base and wide reach. Due to their role in propagating this offence, social media services and messaging services have been included in the risk profiles.

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.

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 could have low financial resilience or be in financial difficulties. 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 on legitimate job sites. This functionality has been included in the risk profiles due to its role in propagating this offence.

Recruiters of money mules[1138] will use U2U services where they can 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 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

6N.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the proceeds of crime 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').

6N.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 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

6N.3 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to 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.

[1138] 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].

6N.5 The priority offences for proceeds of crime are the following:

- Concealing etc. criminal property[1139]
- Arrangements related to criminal property[1140]
- Acquisition, use and possession of criminal property[1141]

6N.6 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

6N.7 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** (Volume 5, Chapter 26).

How proceeds of crime offences manifest online

6N.8 This section is an overview which looks at how proceeds of crime offences manifest online, and how individuals may be at risk of harm.

6N.9 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 and user interactions, in scope of the illegal content safety duties under the OSB. 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".*[1142]

6N.10 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.

6N.11 Criminal organisations often rely on money laundering to conceal the origin of illicit funds. The aim is to make funds appear legitimate and at the same time distance criminal groups from prosecution. This action also makes it difficult for law enforcement to trace money trails. Money mules are people who knowingly or unknowingly help criminal organisations launder their illicit profits. They do this by using their own personal bank accounts to receive and transfer fraudulent funds, thereby making them appear legal.[1143]

6N.12 A variety of scenarios can lead to proceeds of crime offences in the context of user-generated content. A non-exhaustive list of examples is provided below.

- 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

[1139] Section 327 of Proceeds of Crime Act 2002.

[1140] Section 328 of Proceeds of Crime Act 2002.

[1141] Section 329 of Proceeds of Crime Act 2002.

[1142] 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.

[1143] 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 hand over their account details.[1144]

- In other circumstances, mules may be aware or complicit to varying degrees that they are potentially engaging in illegal activity – actively responding to opportunities to make money, shared via user-generated content. These 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 user-generated content. This may include criminals coordinating/arranging money laundering via private messaging or using UGC posts to promote the sale of stolen financial/personal credentials which can be used to launder funds.

6N.13 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".[1145]

6N.14 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.[1146] 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 user-generated content may not be a feature in all cases, social media has been identified as a "*key enabler in the recruitment of mules*".[1147]

6N.15 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.

6N.16 In addition to money muling, the theft and dissemination of stolen credentials online is widespread. 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, as well as fraud 'how to' guides and even fake passports made to order".[1148] The Crown Prosecution Service (CPS), 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.[1149] Cifas has also flagged that identity fraud cases have reached "an all-time high as the cost-of-living crisis bites".[1150]

[1144] Triodos Bank, n.d. What is money muling? [accessed 22 September].

[1145] Sanction Scanner, n.d. The Change of Money Laundering in The Digital Age. [accessed 22 September 2023].

[1146] FBI, n.d. Money Mules. [accessed 22 September 2023].

[1147] Cifas, 2023. Fraudscape 2023. [accessed 22 September 2023].

[1148] Which?, 2020. Your life for sale: stolen bank details and fake passports advertised on social media [accessed 22 September 2023].

[1149] CPS, 2019. Cybercrime - prosecution guidance. [accessed 22 September 2023].

[1150] Cifas, 2023. Identity fraud cases reach all-time high as cost-of-living crisis bites. [accessed 22 September 2023].

Risks of harm to individuals presented by online proceeds of crime offences

6N.17 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[1151] and even suicide.[1152]

6N.18 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.[1153]

6N.19 Ofcom research shows that while anyone who is online is a possible victim, some may be more vulnerable than others.

6N.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.[1154] Cifas identified the key age range for mule activity as 21 to 25,[1155] which was the group hardest hit by the economic impact of Covid-19, with thousands facing job losses as a result of the pandemic, and graduates entering the job market at a time of unprecedented uncertainty.[1156]

6N.21 Adverse economic conditions may increase the likelihood that people will fall victim, whatever their age or income range. Lloyds Bank found that there was a 29% increase in people aged over 40 being used as money mules between October 2021 and September 2022.[1157]

Evidence of risk factors on user-to-user services

6N.22 We consider that the risk factors below are liable to increase the risks of harm relating to proceeds of crime. This is also summarised in the grey box at the start of the chapter.

6N.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, private messaging services, and gaming services.

[1151] 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].

[1152] VICE, 2022. The rise of money launderers on Snapchat and Instagram: Crimewave. [accessed 22 September 2023].

[1153] 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].

[1154] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].

[1155] Cifas, 2023. Fraudscape 2023. [accessed 22 September 2023].

[1156] 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].

[1157] Lloyds Bank, 2022. Money mules are getting older - with serious penalties for those caught moving scam cash. [accessed 22 September 2023].

Social media services

6N.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.[1158] Recruiters of money mules can create user profiles on social media services to join user groups that allow them to find targets,[1159] and use information on users' profiles to befriend potential recruits.[1160] Cifas identifies social media as a 'key enabler' in the recruitment of mules.[1161]

6N.25 Money mule recruiters have also been known to post fake job opportunities on social media services, which are shown to target young people,[1162] and to use specific terms to attract social media users.[1163]

Messaging services

6N.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,[1164] with research showing that services with robust encryption are increasingly used by money mule recruiters to avoid detection.[1165] [1166]

Risk factors: User base

User base size

6N.27 Social media services with user bases that form large user communities are particularly attractive to fraudsters including money mule recruiters[1167] as they make it easy for them to reach large numbers of people[1168] at low cost[1169] and with minimal effort.

6N.28 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.

6N.29 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.

[1158] VICE, 2022. The rise of money launderers on Snapchat and Instagram: Crimewave. [accessed 22 September 2023].

[1159] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].

[1160] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].

[1161] Cifas, 2023. Fraudscape 2023. [accessed 22 September 2023].

[1162] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].

[1163] 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.

[1164] Barclays, n.d. Money Mules: Don't be tricked into committing a crime. [accessed 22 September 2023].

[1165] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].

[1166] See risks of harm to individuals presented by online proceeds of crime offences for more information.

[1167] National Crime Agency, 2021. National Strategic Assessment of Serious and Organised Crime. [accessed 22 September 2023].

[1168] Consumers International, 2019. Social Media Scams: Understanding the Consumer Experience to Create a Safer Digital World. [accessed 22 September 2023].

[1169] Federal Trade Commission (Fletcher, E.), 2022. Social Media a gold mine for scammers in 2021. [accessed 22 September 2023].

User base demographics

6N.30 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.

6N.31 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.

6N.32 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 21-25 years old as the key age range for mule activity.[1170] This group was the hardest hit by the economic impact of Covid-19, with thousands facing job losses as a result of the pandemic, and graduates entering the job market at a time of unprecedented uncertainty.[1171] Lloyds Bank reported that students can be particularly vulnerable as they may be seeking extra income, and the Financial Conduct Authority found that younger people are the most likely to have low financial resilience.[1172] [1173] 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"*;[1174] 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.

6N.33 While younger people are the most likely to be involved in money muling[1175] 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*".[1176] [1177]

6N.34 Research has shown that in May 2022, 12.9 million UK adults had low financial resilience – one in four (24%) of all UK adults. These are people who 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

[1170] Cifas, 2023. Fraudscape 2023. [accessed 22 September 2023].

[1171] 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].

[1172] Lloyds Bank, 2022. Money mules are getting older - with serious penalties for those caught moving scam cash. [accessed 22 September 2023].

[1173] Financial Conduct Authority, 2022. Financial lives 2022 survey: insights on vulnerability and financial resilience relevant to the cost of living. [accessed 22 September 2023].

[1174] Lloyds Bank, 2022. Money mules are getting older - with serious penalties for those caught moving scam cash. [accessed 22 September 2023].

[1175] Most mules are recruited between the ages of 17 and 24. Source: National Crime Agency, n.d. Young People. [accessed 22 September 2023].

[1176] 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].

[1177] Lloyds Bank, 2022. Money mules are getting older - with serious penalties for those caught moving scam cash. [accessed 22 September 2023].

in financial difficulty.[1178] 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.[1179]

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

6N.35 The ability to create a fake user profile provides perpetrators with the opportunity to misrepresent themselves and ultimately mask or conceal 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 accessible to individuals to make contact and get involved in their activity under the false promise of 'fast money'.[1180]

6N.36 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*".[1181] These profiles are often designed to attract young people to a luxurious lifestyle.

User profiles

6N.37 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.[1182]

[1178] Financial Conduct Authority, 2022. Financial lives 2022 survey: insights on vulnerability and financial resilience relevant to the cost of living. [accessed 22 September 2023].

[1179] VICE, 2022. The rise of money launderers on Snapchat and Instagram: Crimewave. [accessed 22 September 2023].

[1180] 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].

[1181] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 22 September 2023].

[1182] Barclays, n.d. Money Mules: Don't be tricked into committing a crime. [accessed 24 August 2023].

User communication

Direct messaging, commenting on content, and re-posting or forwarding

6N.38 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.[1183]

6N.39 There is anecdotal evidence that money mule recruiters may often adopt particular types of shorthand or slang to recruit victims via direct messaging.

Transactions and offers

Post goods or services for sale

6N.40 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*".[1184] Online marketplaces typically allow users to post goods or services for sale.

6N.41 Money mule recruiters may use job websites and social media services to post fake investment and job opportunities. Research from Cifas and UK Finance says that this tactic is particularly targeted towards young people whose job prospects have been damaged by the pandemic.[1185]

Content exploring

User-generated content searching

6N.42 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.[1186] Presumably a means of finding such content is through the search function on user-to-user services that allow for this content to be posted.

Risk factors: Business models and commercial profile

6N.43 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[1183] Keyworth, M., 2018. "I was a teenage 'money mule'". *BBC News,* 26 April. [accessed 22 September 2023].
[1184] CPS, 2019. Cybercrime - prosecution guidance. [accessed 22 September 2023].
[1185] UK Finance, n.d. Money Mule Recruiters Use Fake Online Job Advertys to Target "Generation Covid?. [accessed 22 September 2023].
[1186] Action Fraud, n.d. Money Muling. [accessed 22 September 2023].

6O. 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 misleading statements with the intention to induce users to engage in relevant investment activity.

According to the Office of National Statistics (Crime in England and Wales), fraud is the most frequently experienced crime in the UK. The risks of harms to individuals from fraud and financial services offences is broad. Financial loss to victims is not the only result of fraud; the harm can be multi-faceted and can have a serious impact 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**. Victims of fraud can be targeted through one type of service and will then potentially be moved to another service for further communications and transactions. Different types of fraud may appear on different types of service; for example, investment fraud or ghost broking activities are more likely to occur on social media services. Due to their role in hosting fraud offences, these three types of services are included in the risk profiles.

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; and targeted according to the specific types of fraud. **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, such as 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.

People who have **experienced mental health problems** are also more likely to have been a victim of an online scam.

Low levels of media literacy may also be a key 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 U2U services; these are common across most online services and therefore, in principle, almost any service can be attractive to fraudsters.

The ability to create **fake user profiles** on a U2U service can be used to commit or facilitate fraud, allowing fraudsters 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 a large number of **user connections** to achieve their aims. Due to their role in propagating fraud offences, the functionalities of fake user profiles and user connections are included in the risk profiles.

Functionalities that allow communications between users can be abused by fraudsters. For example, the ability to communicate via **direct messaging or group 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. These messaging functionalities are included in the risk profiles.

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** 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, searching for user-generated content) are both included in the risk profiles for their role in propagating fraud.

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.

Business model risk factors:

Counterfeit goods are often bought on online marketplaces and online auctions that involve **charging a transaction fee.** Boosted posts[1187] on online services offer a gateway for fraudulent actors to access new users, and add legitimacy to the content. The commercial incentive to maximise transaction fees, and boosted posts on services, can provide opportunities to create and promote fraudulent content.

[1187] Boosted posts involve users paying to amplify what would otherwise be a ‘normal’ piece of content on the service.

Introduction

6O.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the fraud and/or financial services 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').

6O.2 We set out the characteristics of user-to-user 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

6O.3 The Online Safety Act (the Act) requires Ofcom to consider the risks of harm connected with specific offences. With regard to 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.

6O.4 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** (Volume 5, Chapter 26).

Priority offences for fraud

6O.5 The priority offences for fraud are the following:

- Fraud by false representation[1188]: 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[1189]: It is an offence to commit fraud by way of a person dishonestly abusing their position
- Making or supplying articles for use in frauds[1190]: 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[1191]
- Participating in fraudulent business carried on by sole trader etc.[1192]: It is an offence for a person to knowingly be a party to the carrying-on of a sole trader etc business with the intent to defraud creditors of any person or for any other fraudulent purpose.

6O.6 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

[1188] Section 2 of the Fraud Act 2006.
[1189] Section 4 of the Fraud Act 2006.
[1190] Section 7 of the Fraud Act 2006.
[1191] Section 49(3) of Criminal Justice and Licensing (Scotland) Act 2010.
[1192] Section 9 of the Fraud Act 2006.

Priority offences for financial services

6O.7 The priority offences for financial services are the following:[1193]

- Contravention of the prohibition on carrying on regulated activity in the UK unless authorised or exempt[1194]
- Falsely claiming to be authorised or exempt for the purposes of carrying-on regulated activity[1195] and the contravention of restrictions on financial promotions[1196]
- Making false or misleading statements, or creating false or misleading impressions about relevant investments[1197]

6O.8 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of these offences.

How fraud and financial services offences manifest online

6O.9 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

6O.10 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 (e.g. a financial institution, the NHS, lottery institution, solicitors, government officials or police officers) and request a payment or information from an individual, potentially via phishing. [1198]
- **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'[1199] 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 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

[1193] These offences are set out in the Financial Services and Markets Act (FMSA) 2000 and the Financial Services Act 2012.

[1194] Section 23 of the FMSA 2000 Act.

[1195] Section 24 of the FMSA 2000 Act.

[1196] Section 25 of the FMSA 2000 Act.

[1197] Section 89 misleading statements or Section 90 misleading impressions of the Financial Services Act 2012.

[1198] "*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].

[1199] 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.

example by offering cryptocurrency. Pension scams are considered as a type of investment scam in this chapter.[1200]

- **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**[1201] seeking to recruit **money mules**[1202] may also commit fraud by false representation via user-generated content in order 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.
- User-generated content can also be used to supply **stolen identity or banking credentials,**[1203] such as stolen identities, credit card details, or 'how to' guides and fake passports.

6O.11 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 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, etc.

6O.12 According to the Crime Survey for England and Wales, fraud, is the most frequently experienced crime in the UK.[1204] 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.[1205]

[1200] *"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, e.g. 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].

[1201] 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.

[1202] 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.

[1203] 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].

[1204] Office of National Statistics (ONS), 2023, Crime in England and Wales: year ending March 2023 [accessed 2 October 2023].

[1205] Ofcom, 2023. Online Scams & Fraud Research [accessed 2 August 2023].

6O.13 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.[1206] 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, please refer to the **Illegal Content Judgements Guidance** (Volume 5, Chapter 26).

6O.14 Ofcom research found that of the 11 examples of scams or fraud tested during quantitative research, *"impersonation fraud (51%) was the most common type that had ever 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%)*".[1207]

6O.15 Ofcom's Online Experiences Tracker (OET) found that generally, the harms eliciting the highest level of expressed concern were encountered at a relatively low claimed incidence. However, '*scams, fraud and phishing*' is an exception, with relatively high levels of concern (74%) and experience (28%), making it the most commonly-experienced potential online harm at the time of fieldwork.[1208]

6O.16 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 2021-22. It was identified that 80% of reported fraud is cyber-enabled and that *"social media and encrypted messaging services as an enabler is increasing throughout all aspects of fraud."*[1209] Money mules are also noted as a persistent feature across most fraud types (see chapter 6N: Proceeds of Crime). 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.

6O.17 While it is challenging to estimate the economic and social cost of fraud, the UK Government Fraud Strategy estimates that the total economic and social cost of fraud to individuals is estimated to be £6.8bn (2019/20).[1210] The UK government estimated the economic and social cost per fraud incident to be £1,427.[1211]

6O.18 Fraud types can overlap and use similar or the same tools to trick victims and gain their trust.[1212] 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. Ofcom research found that of those who had experienced scams or fraud, two in five (41%) said their last experience involved more than one type of scams or fraud.[1213]

[1206] Section 2 of the Fraud Act 2006 [accessed 2 August 2023].

[1207] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1208] Note: Q7 'Please tell me on a scale of 1 to 5, where 1 means 'mildly concerned' and 5 means 'very concerned', how concerned you are about the below existing online.' Was only asked in the first wave of the OET. Source: Ofcom, 2022. The Online Experiences Tracker (2021/22): Summary Report. [accessed 4 September 2023].

[1209] Action Fraud, n.d. Fraud Crime Trends. [accessed 5 September 2023].

[1210] Home Office, 2023. Fraud Strategy: Stopping Scams and Protecting the Public. [accessed 23 August 2023].

[1211] Department for Digital, Culture, Media & Sport, 2022. Online Safety Bill - Impact Assessment. [accessed 4 September].

[1212] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1213] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

Financial services offences: Examples

6O.19 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, potentially in order to defraud victims.

6O.20 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 under the examples above.

6O.21 Consumers have increasingly been exposed to risk via unlawful financial promotions on services such as social media services.[1214] The FCA issued 1,882 alerts[1215] relating to unauthorised activity on its Warning List in 2022, up by 34% from 1,410 in 2021.[1216]

6O.22 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. While user-generated content may not be a feature in all cases, social media was identified as a "*key enabler in the recruitment of mules*",[1217] and of interest to mule herders.

6O.23 For more details on the financial services offences and how services can assess whether content amounts to illegal content, please refer to the **Illegal Content Judgements Guidance** (Volume 5, Chapter 26).

Risks of harm presented by fraud and financial services offences

6O.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 have an impact on 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 a third (34%) reported that the experience had had an immediate negative effect on their mental health.[1218]

6O.25 A report from Action Fraud said that reported losses due to fraud in 2020-2021 amounted to £2.35bn. It also found that adults aged 20-29 experience the greatest volumes of fraud but the greatest losses due to fraud are experienced by 50-69-year-olds.[1219]

6O.26 A qualitative study by Ofcom showed that the impact of a scam continues 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

[1214] DRCF, 2023. 2023/24 Workplan. [accessed 1 August 2023].

[1215] 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. Source: Financial Conduct Authority, 2023. FCA Warning List of unauthorised firms. [accessed 1 August 2023].

[1216] Financial Conduct Authority, 2022. Financial promotions data 2022. [accessed 1 August 2023].

[1217] Cifas, 2023. Fraudscape 2023. [accessed 1 August 2023].

[1218] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1219] Action Fraud, 2021. Fraud Crime Trends. [accessed 1 August 2023].

content and people online afterwards. They can lose confidence in their decision-making, feeling disappointed in themselves and become less trusting of other individuals.[1220]

6O.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, means that they have to pay more for certain goods.[1221]

6O.28 The risks of harm to individuals from financial service offences can also be varied. People who are recruited to be money mules are very often unaware of the consequences and ultimately become victims. A Crimewave video on the NCA website looked at the rise of money laundering on social media sites. It showed that mules who have been recruited sometimes suffer severe effects such as losing their homes and livelihoods, and this can even lead to suicide.[1222] The consequences can also include loss of access to financial services.[1223]

Evidence of risk factors on user-to-user services

6O.29 We consider that the risk factors below are liable to increase the risks of harm relating to fraud and financial services offences. This is also summarised in the grey box at the start of the chapter.

Risk factors: Service types

6O.30 Research indicates that fraudsters use a broad range of types of user-to-user services. These include social media services, messaging services, marketplaces and listings services, and dating services. The majority 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

6O.31 Our evidence reveals that social media services can be used by fraudulent actors in various ways. Ofcom research found that 15% of users' most recent experiences of **impersonation fraud** online originated on social media services and 28% of counterfeit goods scams online were first encountered on social media services.[1224] In 2021, Action Fraud found that during a 12-month period, 5,039 reports of investment fraud made reference to a social media service.[1225] Social media services can also be used to bolster the perceived legitimacy of

[1220] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].
[1221] Age UK, 2023. Age UK- written evidence (DCL0049). [accessed 1 August 2023].
[1222] VICE, 2022. The Rise of Money Launderers on Snapchat and Instagram, *YouTube*, 25 October. [accessed 2 August 2023].
[1223] *"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 1 August 2023].
[1224], 2023. Online Scams & Fraud Research: Data Tables. [accessed 2 August 2023].
[1225] Action Fraud, 2021. New figures reveal victims lost over £63m to investment fraud scams on social media. [accessed 1 August 2023].

fraudsters,[1226] and alongside dating services, are the key enablers of romance fraud where scammers 'love bomb'[1227] their victims[1228]

6O.32 According to the National Fraud Intelligence Bureau (NFIB), social media services or 'social networking sites' and online dating services are the key enablers of **romance scams**.[1229] 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.

6O.33 Instances of 'cloned company investment fraud'[1230] and the use of social media in investment fraud[1231] has increased. In response to our 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.[1232]

6O.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 key enabler of all frauds, and advertising via search engine optimisation[1233] as another source of threat resulting in fraud and scams.[1234]

6O.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**[1235] and **romance fraud**.[1236]

6O.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".* [1237]

[1226] National Fraud Intelligence Bureau 2021

[1227] *"A romantic partner showers you with attention, money, and gifts in order to gain control in a relationship".* Source: Nelson, B., 2022. Is Love Bombing the Newest Scam to Avoid?, *Reader's Digest*, 9 November. [accessed 1 August 2023].

[1228] NFIB Annual Assessment; Source: Clark, J. and Wood, Z., 2023. Victims speak out over 'tsunami' of fraud on Instagram, Facebook and Whatsapp, *The Guardian*, 16 June. [accessed 1 August 2023].

[1229] NFIB Annual Assessment

[1230] FCA, 2021. FCA issues warning over 'clone firm' investment scams [accessed 2 October 2023].

[1231] Investment fraud can relate to financial services offences.

[1232] City of London Police response to 2022 Call for Evidence: First phase of online safety regulation. https://www.ofcom.org.uk/consultations-and-statements/category-1/online-safety-regulation-first-phase

[1233] 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. Source: Search Engine Land, n.d. What Is SEO – Search Engine Optimization?. [accessed 1 August 2023].

[1234] 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".* https://www.ofcom.org.uk/consultations-and-statements/category-1/online-safety-regulation-first-phase

[1235] Identity theft relates to fraud by false representation.

[1236] 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].

[1237] *Which?*, 2020. Your life for sale: stolen bank details and fake passports advertised on social media. [accessed 13 September 2023].

Messaging services

6O.37 Fraudsters use messaging services with encryption to avoid detection and moderation.[1238] They often move conversations, started on other services, to messaging services with encryption.[1239] The increased level of privacy and few verification methods make encrypted messaging services attractive to fraudsters.

6O.38 Research shows that functionalities that are central to many messaging services, such as direct messaging and group messaging, are risk factors for fraud.[1240] Ofcom research found 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%).[1241] The NFIB also found that fraudsters may also use group messaging to store information and to communicate with victims in a group.[1242]

Marketplaces and listings services

6O.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.[1243] 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 platform such as an auction website or social media.*"[1244]

Dating services

6O.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.[1245]

Risk factors: User base

User base size

Services with a large user base

6O.41 Online services with a large user base are particularly attractive[1246] to fraudsters as they make it easy for them to reach large numbers of people at low cost[1247] and with minimal effort.

6O.42 In addition, a large user base makes it more likely that the initial reach of fraudulent UGC posts will be amplified to an even bigger potential audience via a higher volume of content reactions, posts and re-posts.

[1238] National Fraud Intelligence Bureau (2021)

[1239] 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 the National Fraud Intelligence Bureau.

[1240] This may be relevant to offences for fraud and also for financial services.

[1241] Ofcom, 2023. Online Scams and Fraud Research 2022: Data Tables. [accessed 2 August 2023].

[1242] National Fraud Intelligence Bureau (2021)

[1243] Ofcom., 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1244] UK Finance, 2022. Annual Fraud Report. [accessed 1 August 2023].

[1245] 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.

[1246] Consumers International, 2019. Social Media Scams: Understanding the Consumer Experience to Create a Safer Digital World. [accessed 1 August 2023].

[1247] Federal Trade Commission (Fletcher, E.), 2022. Social media a gold mine for scammers in 2021. [accessed 1 August 2023].

6O.43 Fraudsters make use of large, open user groups to add authenticity and to look for potential targets.

6O.44 Having a large number of 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

6O.45 While larger services are a particular target for fraudsters, services with small user bases may also be targeted, depending on the type of fraud. The NFIB has found that some fraudsters look for more niche services in the UK, if these are widely used by certain communities or professions which they can target.[1248]

6O.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

6O.47 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.

6O.48 The data suggests that user base characteristics, including **age**, **financial resilience**, **mental health**, and **media literacy** could lead to an increased risk of harm to individuals for some, or all, types of scam.

6O.49 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.[1249]

6O.50 Users of all ages can be victims of fraud, although different types of scam, and the related content, affect different age groups differently.

6O.51 Investment scams offering returns over time may gain more 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.[1250]

6O.52 The Phoenix Group (a long-term savings and retirement business based in the UK) found that "three in ten 18-34-year-olds fell victim to scams in the past year, with scammers turning to social media to target younger generations".[1251] Younger individuals are the most likely to have low financial resilience, i.e. those aged 18-54 (29%) compared with the UK average of 24%.[1252] This may also mean they are susceptible to the types of fraud that appeal most

[1248] National Fraud Intelligence Bureau (2021).

[1249] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1250] Financial Conduct Authority, 2016. Over 55s at heightened risk of fraud. [accessed 1 August 2023].

[1251] 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].

[1252] *"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* 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.

6O.53 Indeed, the NFIB hypothesised that fraudsters were creating investment fraud opportunities online tailored to low-capital individuals, and creating fraudulent offers that are tempting and accessible to a range of users with differing economic backgrounds but who all use the same online services.[1253]

6O.54 CIFAS identified ages 21-30 as the "*key age range for mule activity*".[1254] Individuals in this age group were the most impacted by the economic impact of Covid-19, with thousands facing job losses as a result of the pandemic, and graduates entering the job market at a time of unprecedented uncertainty.[1255]

6O.55 The FCA's Financial Lives Survey findings suggest that individuals from minority ethnic groups are more likely to have lower financial resilience,[1256] which may mean they are more susceptible to certain fraud types.

6O.56 In response to our call for evidence, the City of London Police said that social media has allowed the victim pool of fraud to become younger, with young people now losing money at a higher rate than older people, by investment and online shopping frauds. Younger people have spent a large proportion of their lives communicating online, so although they may be more tech savvy than older people, they are more open and trusting when sharing personal information in this space.[1257]

6O.57 Mental health problems 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% vs 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*".[1258]

6O.58 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.[1259]

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)". Source: Financial Conduct Authority, 2022. Financial Lives 2022 survey: insights on vulnerability and financial resilience relevant to the rising cost of living. [accessed 1 August 2023].

[1253] National Fraud Intelligence Bureau (2021).

[1254] Cifas, 2023. Fraudscape 2023. [accessed 1 August 2023].

[1255] Cifas, 2021. Money mule recruiters use fake online job adverts to target 'Generation Covid'. [accessed 1 August 2023].

[1256] This is compared with the UK average (Ethnicity: Black & Black British 44%, Mixed/Multiple 39%, UK average 24%). Source: Financial Conduct Authority, 2022. Financial Lives 2022 survey: 3. Low financial resilience. [accessed 1 August 2023].

[1257] City of London Police response to 2022 Call for Evidence: First phase of online safety regulation.

[1258] The Money and Mental Health Policy Institute (Holkar, M. and Lees, C.), 2020. Caught in the Web. [accessed 1 August 2023].

[1259] 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.

6O.59 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 or not. Looking at account verification, 28% of respondents to Ofcom-commissioned research stated that they always, mostly, or sometimes checked for verification symbols when deciding to follow or interact with an account.[1260]

Risk factors: Functionalities and recommender systems

User identification

User profiles

6O.60 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.

6O.61 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 National Fraud Intelligence Bureau (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.[1261] For instance, job and career histories on user profiles can help fraudsters to identify high net-worth individuals.[1262]

6O.62 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 and / 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.[1263]

Fake user profiles

6O.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, a number of public figures co-signed a letter in 2021 discussing how their images had been used by fraudsters to exploit victims' trust.[1264]

6O.64 The fraud prevention service Cifas stated that a key 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. It said: "*this enables criminals to hide behind the anonymity of fake profiles when targeting victims, and to impersonate trusted sources".*[1265]

[1260] Ofcom, 2023. Open Data. [accessed 5 September 2023].

[1261] National Fraud Intelligence Bureau (2021).

[1262] National Fraud Intelligence Bureau (2021).

[1263] Which? (Lipson, F.), 2020. Your life for sale: stolen bank details and fake passports advertised on social media. [accessed 13 September 2023].

[1264] 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].

[1265] Cifas response to 2022 Call for Evidence: First phase of online safety regulation. https://www.ofcom.org.uk/consultations-and-statements/category-1/online-safety-regulation-first-phase

6O.65 Clean Up the Internet[1266] has reported on "*how fraudsters exploit fake social media accounts to scam UK users*".[1267] 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*".[1268]

6O.66 Romance fraudsters use the guise of genuine relationships to manipulate victims for financial gain through creating fake personas on social media services**.** In its response to our call for evidence, the City of London Police said that "s*ocial 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 moving conversations to less regulated and encrypted messaging services*".[1269]

User networking

User connections

6O.67 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, which adds authenticity to their online presence.

6O.68 As a result, user connections can help bolster the perceived legitimacy of fraudsters and their content.[1270] Fraudsters have also used notable entities or 'influencers'[1271] with significant user connections to support their fraudulent work. 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'[1272] 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.[1273]

User groups

6O.69 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.[1274] User groups also make it easier for fraudsters to identify and connect with potential targets.

6O.70 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.[1275]

[1266] Clean Up the Internet is an independent, UK-based organisation concerned about the degradation in online discourse and its implications for society and democracy.

[1267] Babbs, D, 2023. New report on fraud, fake accounts, and the User Verification Duty, *Clean Up the Interne*t, 25 April. [accessed 1 August 2023].

[1268] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1269] City of London Police response to 2022 Call for Evidence.

[1270] National Fraud Intelligence Bureau (2021).

[1271] Influencers gather large followings of enthusiastic, engaged people who pay close attention to their views.

[1272] National Fraud Intelligence Bureau (2021).

[1273] National Fraud Intelligence Bureau (2021).

[1274] 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].

[1275] Which? (Lipson, F.), 2020. Your life for sale: stolen bank details and fake passports advertised on social media. [accessed 13 September 2023].

User communications

Livestreaming

6O.71 Sextortion can be carried out over livestreams.[1276] Sometimes perpetrators use 'fake identities' in these livestreams, presumably by creating inauthentic user profiles, which can be regarded as **fraud by misrepresentation**. 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."*[1277] 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

6O.72 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 this is done through direct messaging (41%).[1278] Depending on their selected settings, users can receive messages from others they may not know, which can lead to a risk of harm.

6O.73 This same research also found that fraudsters often employed several techniques in order 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 (i.e. to get this price you need to sign up in the next 24hrs)".*[1279]

6O.74 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.[1280] 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

6O.75 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.[1281]

[1276] 'Sextortion' is a form of blackmail that involves threatening to publish sexual information, photos or videos about someone.

[1277] National Crime Agency, n.d. Kidnap and Extortion. [accessed 1 August 2023].

[1278] Ofcom, 2023. Online Scams and Fraud Research 2022. [accessed 2 August 2023].

[1279] Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1280] Okpattah, K., 2021. Social media fraud: the influencers promoting criminal scams. *BBC News*, 16 August. [accessed 13 September 2023].

[1281] National Fraud Intelligence Bureau (2021).

Encrypted messaging

6O.76 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 non-encrypted online spaces.[1282] 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.

6O.77 In response to our 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".[1283]

6O.78 In response to our 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".[1284]

Commenting on content

6O.79 The ability to comment on content may enable fraud, although in some contexts these can also help to flag fraudulent activity to other users.

6O.80 Ofcom research found that 8% of 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.[1285] 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 per cent of those who had encountered scams or frauds said they were suspicious because of comments from other users voicing concerns.[1286]

6O.81 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 forum*". 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.[1287]

[1282] National Fraud Intelligence Bureau (2021).

[1283] City of London Police response to 2022 Call for Evidence.

[1284] City of London Police response to 2022 Call for Evidence.

[1285] Ofcom, 2023. Experiences of using online services. [accessed 2 August 2023].

[1286] Ofcom, 2023. Online Scams and Fraud Research 2022. [accessed 2 August 2023].

[1287] Cifas and Forensic Pathways, 2018. Wolves of the Internet. [Accessed: 13 September 2023]

Transactions and offers

Posting goods or services for sale

6O.82 The ability to post goods and services online has been exploited by fraudsters, and the evidence is associated to fraud offences.

6O.83 Data from UK Finance (a banking and finance trade body) indicates that purchase scams are the most common type of 'authorised push payment' fraud, and that these *"usually involve the victim using an online platform such as an auction website or social media*".[1288] While some services which offer the opportunity to post products for sale have secure payment options, fraudsters may prefer services on which they are able to ask users to authorise a payment from their account to the fraudster.

Content exploring

User-generated content searching

6O.84 The ability to search for user-generated content 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 (see below) to others that the content may be fraudulent.

6O.85 Ofcom research found that 6% 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.[1289] Eight per cent said they had first encountered it when using the search function.[1290]

6O.86 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 key terms.[1291] 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.[1292]

6O.87 Research completed by Ofcom[1293] as well as research by other organisations[1294] 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.[1295] Further desk research suggests that, while it is unlikely to be encountered 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.[1296] Once criminals have acquired access to a package of stolen financial credentials and related personal information these will then typically be used to undertake a wide range of secondary fraud activities. These include card-related fraud (e.g. the fraudulent purchase of goods/services/subscriptions, making payments to 'money mule' accounts to launder the

[1288] UK Finance, 2022. Annual Fraud Report. [accessed 1 August 2023].

[1289] Ofcom, 2023. Experiences of using online services. [accessed 2 August 2023].

[1290] Ofcom, 2023. Experiences of using online services. [accessed 2 August 2023].

[1291] Ofcom, 2023. Prevalence of Potentially Prohibited Items on Search Services. [accessed 21 September 2023].

[1292] Okpattah, K., 2021. Social media fraud: The influencers promoting criminal scams, *BBC News,* 16 August. [accessed 23 August 2023].

[1293] Ofcom, 2023. Prevalence of Potentially Prohibited Items on Search Services. [accessed 21 September 2023].

[1294] Okpattah, K., 2021. Social media fraud: The influencers promoting criminal scams, *BBC News,* 16 August. [accessed 23 August 2023].

[1295] SEON, n.d. What Are Fullz. [accessed 4 September 2023]; Fraud.net, n.d. What is Fullz? [accessed 4 September].

[1296] 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.* [accessed 23 August 2023].

proceeds of crime), or impersonation/identity fraud (e.g. stealing someone's identity to take over or set up new bank accounts, email accounts or social media profiles to support fraudulent loan applications, etc).[1297]

6O.88 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.[1298]

Risk factors: Business models and commercial profiles

Revenue models

Transaction fees

6O.89 A recent Ofcom study shows that counterfeit goods are often bought on online marketplaces and online auctions[1299] which involve charging a transaction fee as a key part of their revenue model. Boosted posts, where users pay to amplify their content, are often a feature of these and social media services, and can provide an opportunity for fraudulent actors to access potential victims by presenting a veneer of legitimacy to the content. The commercial incentive to maximise transaction fees, and the boosted posts facility on services such as these can provide opportunities for fraudsters to create and promote fraudulent content.[1300]

[1297] SEON, n.d. What Are Fullz. [accessed 4 September 2023]; Data Dome, 2023. What are fullz? How do fullz work? [accessed 4 September 2023].

[1298] Cifas and Forensic Pathways, 2018. Wolves of the Internet. [accessed: 28 September 2023]

[1299] *"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".* Source: Ofcom, 2023. Online Scams & Fraud Research. [accessed 2 August 2023].

[1300] 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.

6P. Foreign interference offence

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

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, for example, 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 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 election and creating mistrust in online information.

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. For this reason, social media services have been included in the risk profiles.

There is evidence of FIO and influence operations[1301] occurring across many different service types, using different tactics. These services include information-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. It is likely that more attention and resources have been devoted to studying influence operations on some services than on others.

User base risk factors:

Foreign influence operations have targeted **protected characteristics** such as race and gender. For example, targeting women in power to foster gendered narratives and expectations that undermine their power, increasing the risks of harm to them

[1301] 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 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].)

from foreign influence operations. Evidence also suggests that diaspora groups 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. Both fake user profiles and user connections have been included in the risk profiles.

Services where users can more easily share this content onward, 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 impact. These functionalities include **re-posting** and **forwarding content, encrypted messaging,** and mechanisms for sharing information across services, such as the use of **hyperlinks**. These three functionalities are therefore included in the risk profiles.

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 spread harmful content 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

6P.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

a) content on U2U services that may amount to the FIO listed under 'Relevant offences' below, and
b) the use of these services for the commission and/or facilitation of these offences (collectively, the 'risks of harm').

6P.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.

6P.3 The FIO is a new criminal offence set out in clause 13 of the National Security Act 2023.

6P.4 As this is a new offence, there are no precedents.[1302] 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.

6P.5 Attribution is a key challenge associated with the identification of activities that may constitute a 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.[1303] 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.[1304] Recent research has also revealed evidence of inter-state coordination within state-backed information operations, which may further complicate attribution attempts.[1305]

6P.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

[1302] 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.

[1303] Carnegie Endowment for International Peace (Thomas, E., Thompson, N., and Wanless, A.), 2020. The Challenges of Countering Influence Operations. [accessed 11 September 2023].)

[1304] Carnegie Endowment for International Peace (Thomas, E., Thompson, N., and Wanless, A.), 2020. The Challenges of Countering Influence Operations. [accessed 11 September 2023].)

[1305] 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].

attributed to the sponsoring state power. A notable example of this is the Internet Research Agency's[1306] attacks during the US 2018 midterm elections.[1307]

6P.7 The evidence focuses on confirmed foreign interference and influence operations on U2U. 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

6P.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 include the FIO.

6P.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 interfere with the exercise of public functions or prejudice the safety or interests of the UK.[1308]

6P.10 For the FIO to be committed, three conditions must be met:

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).

6P.11 Misrepresentation and coercion are two key types of conduct covered by the FIO:

a) misrepresentation involves making false or misleading misrepresentations, and includes mispresenting a person's identity and using information which is true but presented in a misleading way; and
b) coercion includes using or threatening the use of violence against a person; damaging, destroying or threatening to damage or destroy a person's property or reputation; causing or threatening to cause financial loss to a person; causing spiritual injury or placing undue spiritual pressure on a person.[1309]

6P.12 The Act also covers inchoate offences, such as attempting, conspiring, encouraging, assisting, or aiding and abetting the commission of the offence.

[1306] The Internet Research Agency is a Russian organisation based in St Petersburg that is 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. Mueller, R. S., 2019. Report On The Investigation Into Russian Interference In Source: The 2016 Presidential Election, Volumes I & II. US Department of Justice Publications and Materials. 47. [accessed 11 September 2023].)

[1307] 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].

[1308] Sections 13 and 14 of the National Security Act 2023.

[1309] Section 15 of the National Security Act 2023.

6P.13 For more information on the offence and how services can assess whether content amounts to illegal content, please refer to the **Illegal Content Judgements Guidance** (Volume 5, Chapter 26).

6P.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.

6P.15 Foreign influence is an area of focus for the UK's security services.[1310] When introducing the offence, the 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, its institutions, political systems and rights, or prejudice the UK's national security. The Offence will also seek to protect diaspora communities from the influence of foreign powers.[1311]

6P.16 The 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,[1312] particularly in elections.

6P.17 There is no 'generic' example of what foreign interference looks like.[1313] We know that operations are often intended to sow discord and frequently target sensitive events like 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 interference operation and from which we draw on in this chapter.

6P.18 There are also some examples of events generally considered to have been subject to foreign influence in the UK. These include the leaking of US-UK trade documents during the 2019 General Election,[1314] [1315] Russian influence operations surrounding the attempted assassination of Sergei Skripal,[1316] a network of People's Republic of China-linked accounts

[1310] 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].)

[1311] 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)

[1312]Home Office, 2023. Foreign interference: National Security Act factsheet. [accessed 25 January 2023].

[1313] In the Explanatory Notes for the National Security Act, the 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. Source: Home Office, 2022. National Security Bill: Explanatory Notes. [accessed 27 September 2023].

[1314] 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].

[1315] Graphika (Nimmo, B.), 2019. 'UK Trade Leaks'. [accessed 12 September 2023].

[1316] Global Engagement Centre, 2022. GEC Special Report: The Kremlin's Chemical Weapons Disinformation Campaigns. [accessed 12 September 2023].

amplifying the activities of Chinese diplomats in the UK online[1317], and Russian influence operations following its invasion of Ukraine in February 2022.[1318]

How foreign interference manifests online

6P.19 This section is an overview which looks at how FIO offences manifest online, and how individuals may be at risks of harm.

6P.20 By their nature, foreign interference activities are typically covert and due to the involvement of nation states, are often sophisticated in nature. Noting the limited evidence to draw robust conclusions at this time, this section is an overview that outlines how the FIO can manifest online, and how individuals may be at risk of harm.

6P.21 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.[1319] There is no robust data about the number of UK internet users exposed to those campaigns, nor the nature of the campaigns' audiences.

6P.22 In the Explanatory Notes for the National Security Act, the 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.[1320]

6P.23 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.

6P.24 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.

[1317] 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].

[1318] Foreign, Development and Commonwealth Office, 2022. UK Exposes Sick Russian Troll Factory plaguing Social Media with Kremlin Propaganda. [accessed 27 September 2023].

[1319] 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].

[1320] Home Office, 2022. National Security Bill: Explanatory Notes. [accessed 27 September 2023].

6P.25 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.[1321] Such campaigns have targeted the UK,[1322] and evidence shows that Russia, the People's Republic of China, Saudi Arabia and the UAE have all engaged in cross-jurisdictional operations.[1323] 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.[1324]

6P.26 The evidence we have assessed suggests that the deployment of bots[1325] 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.

6P.27 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[1326] and click farming,[1327] hashtag hijacking,[1328] initiating a repost storm (when a post is instantly reposted by a network of accounts) and trend-jacking.[1329] [1330] 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.[1331], [1332]

6P.28 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.

[1321] 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].

[1322] Carnigie Endowment For International Peace (Thomas, E., Thompson, N. and Wanless, A.), 2020. The Challenges of Countering Influence Operations. [accessed 27 September 2023].

[1323] 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].

[1324] Threat Analysis Group (Butler, Z. and Taege, J.), 2023. Over 50,000 instances of DRAGONBRIDGE activity disrupted in 2022. [accessed 27 September 2023].

[1325] '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.

[1326] '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.

[1327] '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.

[1328] '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.

[1329] '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.

[1330] US Department of Homeland Security, 2018. Social Media Bots Overview. [accessed 13 September 2023].

[1331] 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].

[1332] It is unclear whether some of these bots employ GenAI technologies, but we think that GenAI bots could be used in a similar manner.

Risks of harm to individuals presented by the foreign interference offence

6P.29 Overall, the evidence available does not allow us to draw robust conclusions about the impact and harm associated with foreign influence operations. However, such operations are 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*".[1333] Despite the lack of direct insight, there is clear potential that there is a risk of harm to individuals which underpins the desire by government to introduce this offence in the online safety regime.

6P.30 Media and public concern over foreign influence operations in the UK has 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.[1334]

Evidence of risk factors on user-to-user services

6P.31 We consider that the risk factors below are liable to increase the risks of harm relating to FIO. This is also summarised in the grey box at the start of the chapter.

Risk factors: Service types

6P.32 Research shows that a broad range of services can be used to commit or facilitate offences related to foreign interference. In particular, 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.[1335]

Social media services

6P.33 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,[1336] 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

[1333] 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].

[1334] 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].

[1335] 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.

[1336] 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].

Russian disinformation operation in the UK and several other countries, targeting Kremlin critics on social media services.[1337]

Information-sharing services

6P.34 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 consistent with Russian state-sponsored information warfare.[1338] The Wikimedia Foundation has also banned several editors linked to a group from the People's Republic of People's Republic of China.[1339]

Discussion forums and chat rooms

6P.35 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.[1340] 'Online forums' were also used in a North Korean operation in 2010, through fake accounts that disseminated disinformation.[1341.]

6P.36 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

6P.37 Messaging services are also a risk factor for foreign interference. Research shows that private messaging applications are increasingly common channels for foreign influence operations[1342]. The ability to forward content on messaging services has been a key mechanism for the spread of disinformation, including foreign influence operation content.

6P.38 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.[1343] [1344]

[1337] 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].

[1338] 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].

[1339] 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].

[1340] 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].

[1341] 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].

[1342] 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].

[1343] 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].

[1344] Carnegie Endowment for International Peace (Goodwin, C. and Jackson, D.), 2022. Global Perspectives on Influence Operations Investigations: Shared Challenges, Unequal Resources. [accessed 27 September 2023].

Risk factors: User base

User base demographics

6P.39 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.

6P.40 Data suggests that user base characteristics including the **gender and race** of users could lead to an increased risk of harm to individuals.

6P.41 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.[1345]

6P.42 Research has found that political leaders in Russia, the Philippines, Hungary and Turkey have used gendered disinformation campaigns to attack women in politics.[1346] [1347] 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.[1348]

6P.43 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.

6P.44 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.[1349] 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.[1350]

[1345] 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].

[1346] Di Meco, L. and Wilfore, K., 2021. Gendered disinformation is a national security problem, *Brookings Institute*, 8 March. [accessed 6 March 2023].

[1347] 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].

[1348] HM Government Stabilisation Unit, 2020. Quick-read guide: gender and countering disinformation. [accessed 6 March 2023].

[1349] Freedom House (Datt, A. and Dunning, S.), 2022. Beijing's Global Media Influence 2022. [accessed 17 February 2023].

[1350] Hope, C., 2021. Exclusive: Uighurs harassed and abused by Beijing in UK, minister admits, *The Telegraph*, 13 March. [accessed 17 February 2023].

Risk factors: Functionalities and recommender systems

User identification

User profiles

6P.45 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 (e.g. information about language spoken or home town) could also increase the risks of harm to these diaspora groups.

6P.46 There is also some evidence suggesting the targeting of users by political ideology on specific social media services with divisive narratives.[1351]

Fake user profiles

6P.47 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. [1352] The unit claims to use a variety of tactics, including automated disinformation on a range of social media services by creating fake accounts. 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.[1353] Journalists also found evidence that campaigns using this product had previously targeted the UK, as well as other countries.[1354]

6P.48 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.[1355]

6P.49 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.[1356]

6P.50 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.[1357] Additionally, in 2022, Meta

[1351] Graphika, 2020, Step into my Parler. [accessed 22 September 2023].
[1352] 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].
[1353] 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].
[1354] Ganguly, M., 2023. 'Aims': the software for hire that can control 30,000 fake online profiles, *The Guardian*, 15 February. [accessed 15 February 2023].
[1355] TikTok, 2022. Community Guidelines Enforcement Report. accessed 15 February 2023].
[1356] REPORT European Parliament (Kalniete, S.), 2022. REPORT on foreign interference in all democratic processes in the European Union, including disinformation. [accessed 27 September 2023].
[1357] Facebook, 2021. January 2021 Coordinated Inauthentic Behaviour Report. [accessed 28 June 2023].

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.[1358] The content created by these sites was amplified on social media by many fake accounts, which often claimed to work for organisations like Netflix.

6P.51 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 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 (e.g. account names containing 'UK').[1359]

Anonymous user profiles

6P.52 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.[1360] For example, it is alleged that anonymous users on a service were involved in foreign influence operations during the 2017 French Presidential election.[1361]

6P.53 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.[1362]

User networking

User connections

6P.54 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 example below) to wider audiences and can be used to spread narratives across social media platforms through the simultaneous posting of similar or identical content.

[1358] Meta, 2022. Removing Coordinated Inauthentic Behavior From People's Republic of China and Russia. [accessed 27 September 2023].

[1359] 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].

[1360] In such cases, the identity of users may also be unknown to services.

[1361] 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].

[1362] 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].

6P.55 For example, the Oxford Internet Institute identified a network of 62 Twitter accounts that amplified and engaged with UK-based diplomats from the People's Republic of 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.[1363]

User groups

6P.56 The evidence we have assessed suggests that the creation and use of groups can be exploited by perpetrators of foreign influence campaigns.

6P.57 It is well evidenced that services with group functionalities can instigate users to conduct offline activity,[1364] 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.[1365]

6P.58 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.[1366] 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 which targeted people in Ukraine, promoting claims that the West had betrayed Ukraine and that Ukraine was a failed state.[1367]

User tagging

6P.59 The evidence suggests that the ability to tag users can be exploited by perpetrators of foreign interference campaigns.

6P.60 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;[1368] Reddit attributed a network of 61 accounts sharing these documents to Russia.[1369] A Twitter account was used to promote links to the leaked documents by tagging opposition politicians and prominent journalists.[1370]

[1363] 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].

[1364] Thiel, D. and McCain, M. 2022. Gabufacuturing Dissent: An in-depth analysis of Gab, Stanford Cyber Policy Review. [accessed 27 September 2023].

[1365] 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].

[1366] Meta, 2022. January 2022 Coordinated Inauthentic Behaviour Report. [accessed 17 February 2023].

[1367] Meta (Nimmo, B., Agranovich, D. and Gleicher, N.), 2022. Adversarial Threat Report. [accessed 17 February 2023].

[1368] Graphika (Nimmo, B.), 2019. UK Trade Leaks. [accessed 27 September 2023].

[1369] 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].

[1370] Graphika (Nimmo, B.), 2019. UK Trade Leaks. [accessed 27 September 2023].

User communication

Direct messaging, group messaging, encrypted messaging

6P.61 The creation and amplification of disinformation content is a key 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 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.[1371] 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.[1372]

Reacting to content

6P.62 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.[1373]

Posting content

6P.63 **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.[1374] Other types of multimodal disinformation are outlined in the 'content editing' section below.

6P.64 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.[1375]

[1371] Wardle, C., 2018: 5 Lessons for Reporting in an Age of Disinformation, *First Draft News*, 27 December. [accessed 21 September 2023].

[1372] 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].

[1373] Koval, I., 2021. "How social media is manipulated — and how Russia is involved", *DW*, 14 April, [accessed 22 June 2023].

[1374] 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].

[1375] 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].

Re-posting or forwarding content

6P.65 Frequently forwarded and mass-forwarded messages, sent directly from a trusted set of contacts on an encrypted private messaging service to more open services,[1376] have been a key 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.[1377]

6P.66 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.[1378]

6P.67 The 'doxing' of individuals is another tactic used within foreign influence operations. Doxing 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 doxed 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.[1379]

Content exploring

Hyperlinking

6P.68 Many foreign influence operations are run concurrently across different platforms. 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 IRA-linked accounts into their reporting.[1380]

6P.69 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 content gained illicitly through hack and leak operations, and by amplifying doxing campaigns[1381].

6P.70 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

[1376] 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].

[1377] WhatsApp, n.d. About forwarding limits | WhatsApp Help Center. [accessed 27 September 2023].

[1378] 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].

[1380] 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].

[1381] Institute for Strategic Dialogue (Thomas, E.), 2022. Project Nemesis, Doxxing and the New Frontier of Informational Warfare. [accessed 27 September 2023].

YouTube are driven by external referral sites like Gab. Other commonly-linked-to sites include Facebook, Twitter and closed Telegram groups.[1382]

6P.71 '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-UK trade documents, to Russia.[1383] A similar operation is reported to have been used to target and dox allegedly British-influenced individuals in Russia on social media.[1384]

Content editing

Editing visual media

6P.72 Cheap-fakes[1385] and deepfakes[1386] have been a key concern for information operations researchers since they came to prominence in 2017.[1387] 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 than they contain in their non-edited form.[1388] These are different types of **multimodal** disinformation (see 'posting content' for more).

6P.73 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.[1389] Another example is a low-quality deepfake of Ukrainian President Volodymyr Zelenskyy talking about surrendering to Russia on social media services.[1390] Although the video has not been specifically attributed to a state actor, it is widely believed that it was generated by Russia.

6P.74 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 fictious people being promoted by Spamouflage, a pro-Chinese influence operation.[1391]

[1382] Thiel, D. and McCain, M., 2022. Gabufacuturing Dissent: An in-depth analysis of Gab, Stanford Cyber Policy Review. [accessed 27 September 2023].

[1383] Wendling, M., 2019. General election 2019: Reddit says UK-US trade talks document leak 'linked to Russia', *BBC*, 7 December. [accessed 27 September 2023].

[1384] Institute for Strategic Dialogue, 2022. Tales From the Underside: A Kremlin-Approved Hack, Leak & Doxxing Operation. [accessed 27 September 2023].

[1385] Cheap-fakes are videos that use conventional video editing techniques like speeding, slowing, cutting, restaging or re-contextualising video footage

[1386] 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.

[1387] Donovan, J. and Paris, B., 2019. Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence, *Data & Society.* [accessed 27 September 2023].

[1388] 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].

[1389] Donovan, J. and Paris, B., 2019. Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence, *Data & Society.* [accessed 27 September 2023].

[1390] Wakefield, J., 2022. Deepfake presidents used in Russia-Ukraine war, *BBC News*, 18 March. [accessed 6 March 2023].

[1391] Graphika, 2023. Deepfake It Till You Make It. [accessed 17 February 2023].

6P.75 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.[1392]

Editing posted content

6P.76 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.[1393]

6P.77 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[1394] which could be vulnerable to influence operations, including foreign influence operations. Other practices include undisclosed paid editing, adversarial editing and state editing.

6P.78 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.[1395]

6P.79 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.

6P.80 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.[1396]

[1392] Thurston, J, 2023. Russia deepfake video mocks Rishi Sunak and Joe Biden, *The Times,* 15 June. [accessed 27 September 2023].
[1393] Robson, K., 2022. Will Twitter's edit button help spread more fake news?, *Verdic*t, 2 September 2022. [accessed 27 September 2023].
[1394] 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].
[1395] 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].
[1396] 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

6P.81 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.[1397] 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.[1398]

Recommender systems

Content recommender systems

6P.82 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).[1399] 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 registered as positive feedback on all types of content). We consider that content recommender systems may be manipulated by perpetrators of foreign influence operations.

[1397] 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].

[1398] 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].

[1399] 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.

Risk factors: Business model and commercial profile

Revenue models

6P.83 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,[1400] 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".*[1401] The report added that if users are unaware of the political intent 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.[1402]

6P.84 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'.[1403]

6P.85 The same study recognises how ad targeting[1404] 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.

6P.86 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.[1405]

[1400] 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."*

[1401] 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].

[1402] 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].

[1403] 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].

[1404] 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.

[1405] Colliver, C., King, J. and Maharasingam-Shah, E., 2020. Hoodwinked: Coordinated Inauthentic Behaviour on Facebook. [accessed 27 September 2023].

6Q. False Communications Offence

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

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-ranging. 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, as could occur, for instance, from hoax bomb threats.

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, which could apply to the false communications offence. 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:

The **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:

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 key 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 by creating deepfakes, can also be exploited by perpetrators of the false communications offence.

Introduction

6Q.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

- content on U2U services that may amount to the false communications 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').

6Q.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.

Relevant offences

6Q.3 In this chapter we consider the new offence of sending false communications, established by section 179 of the Online Safety Act (the Act). It is an offence for a person to send a message that they know to be false, if in sending that message they intend to cause non-trivial psychological or physical harm to a likely audience, and they have no reasonable excuse for sending the message.[1406]

6Q.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.[1407]

6Q.5 The Act also covers the offences of encouraging and assisting, and conspiracy to commit, this offence.

6Q.6 For more details on the offence and how services can assess whether content amounts to illegal content, please refer to the **Illegal Content Judgements Guidance** (Volume 5, Chapter 26).

How false communication manifests online

6Q.7 This section is an overview which looks at how the false communications offence manifests online, and how individuals may be at risk of harm.

6Q.8 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.[1408] It aims to cover "*false communications deliberately sent to inflict harm,*

[1406] The Act provides a definition for 'likely audience' at section 179(2) as well as a number of exemptions for recognised news publishers and broadcasters (section 180).

[1407] UK Government (Department for Digital, Culture, Media and Sport), 2022. Online safety law to be strengthened to stamp out illegal content.

[1408] 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.

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".[1409]

6Q.9 As this is a new offence, in preparing this risk assessment we have considered evidence of conduct that appears to be broadly aligned with the conduct that is 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 new offence. We will keep our evidence base under review as new evidence emerges, once the offence is law.

Risks of harm to individuals presented by the false communication offence

6Q.10 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 (chapter 6P: Foreign interference offence) and with some of the offences discussed in the chapter on fraud (chapter 6O: Fraud and financial services offences).

6Q.11 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:

- De-contextualisation: When real images or videos are paired with false or manipulated text.
- Reframing: When videos are cropped or decontextualised to make certain aspects or issues more obvious or prominent in pursuit of a specific agenda.
- Visual doctoring: When images or videos are manipulated to present a different reality than they contain in their non-edited form – this covers both cheap-fakes and deepfakes.
- Multimodal doctoring: When content is fabricated by pairing manipulated images or videos with false or manipulated content.[1410]

6Q.12 The evidence we have assessed suggests that the deployment of bots[1411] 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.

[1409] 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].

[1410] 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)

[1411] '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.

6Q.13 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[1412] and click-farming[1413], hashtag hijacking[1414], initiating a repost storm (when a post is instantly reposted by a network of users) and trend-jacking.[1415] [1416] We believe that bots, including those which might employ GenAI technologies, could be used in a similar way by perpetrators of the false communication offence.

6Q.14 We expect that false communications are done to elicit a response from someone. This may be to influence them into undertaking a certain activity or to cause psychological impacts to an individual, such as fear and anxiety.

Evidence of risks of harm on user-to-user services

6Q.15 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

6Q.16 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

6Q.17 Social media services can be used by potential perpetrators to spread disinformation. For example, social media services were used to disseminate false claim 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.[1417] An in-depth investigation led by the European Centre for Disease Prevention and Control looked into a large number of misinformation campaigns surrounding the Covid-19 vaccine and found some were being spread with malicious intent – these could cause harm.[1418] This content could then be picked up and amplified by organisations with a greater potential reach, thereby increasing the number of users who could encounter it (see Posting content section below for more information).

6Q.18 As mentioned in the section above, bots programmed to spread disinformation are also 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.

[1412] '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.

[1413] '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.

[1414] '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.

[1415] '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.

[1416] US Department of Homeland Security, 2018. Social Media Bots Overview.

[1417] Facebook, 2021. July 2021 Coordinated Inauthentic Behavior report. [accessed 1 September 2023].

[1418] European Centre For Disease Prevention And Control, 2021. Countering online vaccine misinformation in the EU/EEA. [accessed 20 September 2023].

Messaging services

6Q.19 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 as a result, "*these platforms have become a promising new avenue for the spread of disinformation, particularly among diaspora communities*" [1419] (see Risk factors: user base section for more information).

Risk factors: user base

User base size

6Q.20 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 chapter 6W: Context to understand risk factor dynamics (in the Annex of this document).

User base demographics

6Q.21 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.

6Q.22 The data suggest that the user base characteristics of **gender, diaspora communities, and religion** could lead to increased risks of harm to individuals.

6Q.23 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. [1420] **Although we are still unsure about conduct that would constitute this offence, it is highly likely the false communications offence may use similar tactics.**

6Q.24 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 Ukrainian army lost an important battle, and doctored images, purporting to show her doing so. [1421]

[1419] Gorksy, J., Riedl,M.J, and Woolley, S., 2021. *The Disinformation Threat to Diaspora Communities in Encrypted Chat Apps*, Tech Stream, Brookings Institute.

[1420] 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].

[1421] Jankowicz, N, 2017. How disinformation became a new threat to women. [accessed 6 March 2023].

6Q.25 During the 2020 US Presidential election, a hoax about 'disregarded ballots' spread online using encrypted private messaging services, particularly among diaspora communities.[1422] It is reasonable to assume that a false communications offence may include the spread of hoaxes across diaspora groups in a similar way.

6Q.26 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 specific religious affiliations, 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.[1423]

Risk factors: Functionalities and recommender systems

User identification

Fake user profiles

6Q.27 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 chapter 6P: Foreign interference offence for more information).

Anonymous user profiles

6Q.28 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.

6Q.29 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.[1424]

[1422] 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].

[1423] Kotecha, S., 2021. *Covid: Fake news 'causing UK South Asians to reject jab, BBC News,* 15 January. [accessed 9 June 2023].

[1424] 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].

User communication

Posting content

6Q.30 The evidence we have assessed suggests that the ability to post content online is key to the false communications offence.

6Q.31 Think-tank and campaign group First Draft's article 'Trumpet of Amplification' highlights how individuals spreading disinformation have used encrypted and anonymous online 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*." [1425]

6Q.32 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. [1426]

Direct messaging and encrypted messaging

6Q.33 Direct messaging represents another means by which false communications can be disseminated. Encrypted messaging can also be used. For instance, our research shows that encrypted private messaging services can be used to spread disinformation, particularly among diaspora communities. [1427] 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

6Q.34 The creation of deepfakes [1428] can be exploited by perpetrators of the false communications offence. [1429] 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. [1430] [1431]

Risk factors: Business models and commercial profiles

6Q.35 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[1425] First Draft, (Wardle, C.), 2018. 5 Lessons for Reporting in an Age of Disinformation. [accessed 20 September 2023].

[1426] Facebook, 2021. July 2021 Coordinated Inauthentic Behaviour Report. [accessed 12 June 2023].

[1427] Gorksy, J., Riedl, M.J, and Woolley, S., 2021. *The Disinformation Threat to Diaspora Communities in Encrypted Chat Apps.* Tech Stream, Brookings Institute.

[1428] 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.

[1429] Donovan, J, and Paris, B., 2019. Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence. [accessed 20 September 2023].

[1430] BBC News, (Wakefield, J.), 2022. Deepfake presidents used in Russia-Ukraine war. 18 March. [accessed 6 March 2023].

[1431] Gleicher, N., 2022. Tweet on removal of Zelenskyy deepfake, published 16 March 2022 [accessed 6 March 2023].

6R. 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.

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 to trigger seizures. **Tagging content** through hashtags, for instance, can also increase the risk of content with the potential to trigger seizures being disseminated on a service. **Recommender systems** can increase the risk that this content, if tagged with epilepsy-related hashtags, will be seen by those engaging with epilepsy-related content (perhaps because they have epilepsy themselves).

Introduction

6R.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

a) content on U2U services that may amount to the epilepsy trolling offence listed under 'Relevant offences' below; and

b) the use of these services for the commission and/or facilitation of this offence (collectively the 'risks of harm'). 1432

6R.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. [1433]

6R.3 This offence is new under the Online Safety Act (the Act), and the Epilepsy Society played a key role in its inclusion. 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

6R.4 In this chapter we consider the offence of sending or showing flashing images electronically, a new offence set out in section 183 of the Act. Commonly referred to as 'epilepsy trolling', this offence involves sending flashing images[1434] electronically to trigger seizures, or cause alarm or distress, among people with epilepsy, in particular those with photosensitive epilepsy.

6R.5 The Act also covers the offences of encouraging and assisting, and conspiracy to commit, this offence.

6R.6 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** (Volume 5, Chapter 26).

How epilepsy trolling manifests online

6R.7 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 offences, recognising that the specific examples given, by definition, pre-date the new offence itself.

[1432] 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.

[1433] We note that the epilepsy offence itself adopts a different definition of harm (see section 183(13)).

[1434] Flashing images include GIFs.

6R.8 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.[1435]

6R.9 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".* [1436]

6R.10 In May 2020 there was a sustained *"attack by internet trolls"*[1437] on the Epilepsy Society's 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'.[1438]

6R.11 To provide context on how many users may have been exposed to this behaviour, the Epilepsy Society Twitter page currently has more than 30,000 followers.[1439] It is unclear how many of these may have had photosensitive epilepsy themselves.

6R.12 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.[1440]

Risks of harm to individuals presented by the offence of epilepsy trolling online

6R.13 In the May 2020 series of attacks involving the Epilepsy Society's Twitter 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.

6R.14 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).[1441]

6R.15 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.[1442] Seizures can also be fatal due to multiple outcomes that may result from them.[1443] This offence may also cause emotional harm by causing alarm or distress among those with epilepsy.

[1435] Epilepsy Action, n.d. Photosensitive epilepsy. [accessed 10 May 2023].

[1436] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].

[1437] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 15 June 2023].

[1438] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].

[1439] Source: Epilepsy Society, n.d. Epilsepsy Society Twitter Page. [accessed 31 July 2023].

[1440] Meeting between Ofcom and Epilepsy Society, April 2023.

[1441] Meeting between Ofcom and Epilepsy Society, April 2023.

[1442] Epilepsy Society, n.d. Photosensitive epilepsy. [accessed 2 May 2023].

[1443] Meeting between Ofcom and Epilepsy Society, April 2023.

6R.16 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. [1444]

6R.17 Written evidence from the Epilepsy Society submitted to Government described two cases of Twitter users who had been harmed, demonstrating that both adults[1445] and children[1446] can be affected by this harm.

Evidence of risk factors on user-to-user services

6R.18 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

6R.19 To date, attacks have taken place on social media services (predominately Twitter).[1447] 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

6R.20 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.

6R.21 Evidence from the Epilepsy Society suggests that user base characteristics including **age** and **disability** could lead to increased risks of harm to an individual.

6R.22 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 sometimes individuals with photosensitive epilepsy will experience fewer symptoms once they enter their mid to late twenties. [1448]

[1444] Meeting between Ofcom and Epilepsy Society, April 2023.

[1445] One example was that of a 25-year-old man who had recently been diagnosed with epilepsy. He visited the Epilepsy Society Twitter 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].

[1446] In May 2020 a young boy aged 8 was harmed when his mum proudly shared a video of him on Twitter 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].

[1447] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].

[1448] Meeting between Ofcom and Epilepsy Society, April 2023.

6R.23 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.[1449]

Risk factors: Functionalities and recommender systems

User identification

User profiles and anonymous user profiles

6R.24 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.[1450]

6R.25 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.[1451]

6R.26 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.[1452]

User networking

User groups

6R.27 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

6R.28 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.[1453]

User tagging

6R.29 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 Epilepsy Society's Twitter 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.[1454]

[1449] Epilepsy Action, n.d. Photosensitive epilepsy. [accessed 10 May 2023].

[1450] Meeting between Ofcom and Epilepsy Society, April 2023.

[1451] Meeting between Ofcom and Epilepsy Society, April 2023.

[1452] 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].

[1453] Meeting between Ofcom and Epilepsy Society, April 2023.

[1454] Meeting between Ofcom and Epilepsy Society, April 2023.

User communications

Direct messaging

6R.30 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.[1455]

Posting content and commenting on content (images and videos)

6R.31 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.[1456]

6R.32 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.[1457]

6R.33 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. [1458]

Content exploring

Content tagging

6R.34 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.[1459]

6R.35 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.[1460]

Hyperlinking

6R.36 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.

[1455] Meeting between Ofcom and Epilepsy Society, April 2023.

[1456] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].

[1457] Meeting between Ofcom and Epilepsy Society, April 2023.

[1458] Epilepsy Society, n.d. Call for evidence on the Draft Online Safety Bill. [accessed 17 April 2023].

[1459] Meeting between Ofcom and Epilepsy Society, April 2023.

[1460] Meeting between Ofcom and Epilepsy Society, April 2023.

Content editing

Editing visual media

6R.37 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 in to a GIF. These would typically be of music videos that featured clips with flashing images, which would then be posted to users with epilepsy.[1461]

6R.38 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. [1462]

Recommender systems

Content recommender systems

6R.39 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.

6R.40 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.[1463]

6R.41 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.[1464]

6R.42 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.[1465]

Risk factors: Business models and commercial profiles

6R.43 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[1461] Meeting between Ofcom and Epilepsy Society, April 2023.

[1462] Meeting between Ofcom and Epilepsy Society, April 2023.

[1463] Meeting between Ofcom and Epilepsy Society, April 2023.

[1464] Meeting between Ofcom and Epilepsy Society, April 2023.

[1465] Meeting between Ofcom and Epilepsy Society, April 2023.

6S. Cyberflashing offence

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.

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.

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; the majority of victims and survivors are women and the majority of the perpetrators are men. 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

6S.1 This chapter summarises our assessment of the risks of harm to individuals presented by:

a) content on U2U services that may amount to the cyberflashing offence listed under 'Relevant offences' below; and
b) the use of these services for the commission and/or facilitation of these offences (collectively the 'risks of harm').

6S.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.

6S.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 6C: Child sexual exploitation and abuse offences.[1466]

6S.4 There are two primary mechanisms used by perpetrators to cyberflash: Bluetooth channels[1467] 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.

6S.5 There is currently limited research and evidence available on cyberflashing via online means, and its impacts. Much of the research is focused on unsolicited sexual images rather than cyberflashing as it is defined in the legislation.

Relevant offences

6S.6 In this chapter we consider the offence of cyberflashing, a new offence[1468] created by the Act.[1469] 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.

6S.7 The Act also covers the offences of encouraging and assisting, and conspiracy to commit, this offence.

6S.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** (Volume 5, Chapter 26).

6S.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 (chapter 6M: intimate image abuse offences), cyberflashing is a form of image-based sexual abuse and is disproportionately perpetrated by men towards women.[1470]

[1466] 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].

[1467] 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.

[1468] New section 66A of the Sexual Offences Act 2003.

[1469] Section 187 of the Online Safety Act 2023.

[1470] Law Commission, 2021. Modernising Communications Offences: A final report. [accessed 26 July 2021].

6S.10 Further, the sending of unsolicited sexual images cannot always be described as cyberflashing, as the intent to harm cannot always be assumed. In some circumstances, those receiving the unsolicited sexual image may view it positively.[1471]

How cyberflashing manifests online

6S.11 This section is an overview which looks at how cyberflashing manifests online, and how individuals may be at risk of harms.

6S.12 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.[1472] Among young people, rates of cyberflashing are even higher[1473]; 47% of women aged 18-24 have been cyberflashed[1474] and 19% of girls aged 11-16 have been sent unwanted sexual images.[1475] Some individuals report very high levels of cyberflashing, receiving several images a day.[1476]

Risks of harm to individuals presented by cyberflashing

6S.13 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.[1477]

6S.14 Cyberflashing is a gendered offence; the majority of victims and survivors are women and the majority of the perpetrators are men.[1478] There is also evidence to suggest that women in minority ethnic groups and LGBT+ groups disproportionately experience cyberflashing.[1479]

6S.15 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.

6S.16 Testimonies from victims and survivors of cyberflashing describe feelings of shame and embarrassment.[1480] Of women who have been cyberflashed, 58% described the unsolicited images as ‘gross’ and 54% described them as ‘stupid’.[1481]

[1471] Marcotte, A., Gesselman, A., Fisher, H. and Garcia, J., 2021. Women’s and Men’s Reactions to Receiving Unsolicited Genital Images From Men, *The Journal of Sex Research, 58* (4). [accessed 5 September 2023].

[1472] 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].

[1473] 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].

[1474] YouGov (Smith, M.), 2018. Four in ten female millennials have been sent an unsolicited penis photo. [accessed 26 July 2023].

[1475] Sample consisted of 2,114 girls and young women aged 7-21. Source: Girlguiding, 2021. Girls’ Attitudes Survey. [accessed 26 July 2023].

[1476] 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].

[1477] 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.

[1478] Law Commission, 2021. Modernising Communications Offences: A final report. [accessed 26 July 2023].

[1479] 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].

[1480] Law Commission, 2021. Modernising Communications Offences: A final report. [accessed 26 July 2021].

[1481] 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].

6S.17 Victims and survivors also describe feeling vulnerable following experiences of cyberflashing.[1482] Many victims and survivors describe the experience of being cyberflashed as aggressive and intimating,[1483] and some victims and survivors describe the experience as a violation.[1484]

6S.18 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.[1485]

6S.19 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.[1486] As with all forms of sexual violence, perpetrators of this abuse are motived by a desire to exert power,[1487] and victims and survivors experience feelings of fright and vulnerability.[1488]

6S.20 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,[1489] harassment, and controlling or coercive behaviour.[1490]

Evidence of risk factors on user-to-user services

6S.21 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

6S.22 Cyberflashing offences can occur on any service that enables users to share images. Evidence suggests that social media and online dating services are particularly risky service types for cyberflashing offences.[1491]

Social media services

6S.23 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.[1492]

[1482] Law Commission, 2021. Modernising Communications Offences: A final report. [accessed 26 July 2021].
[1483] The National Police Chiefs' Council response to 2020 Law Commission Consultation. [accessed 2 August 2023].
[1484] McGlynn, C. response to 2020 Law Commission Consultation. [accessed 2 August 2023].
[1485] McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. Bristol: Bristol University Press. [accessed 21 September 2023].
[1486] McGlynn, C., 2022. Cyberflashing: Consent, Reform and the Criminal Law, *The Journal of Criminal Law,* 85 (5). [accessed 28 July 2023].
[1487] McGlynn, C., 2021. Written evidence submitted by Professor Clare McGlynn, Durham Law School, Durham University (VAW0007). [accessed 28 July 2023].
[1488] Law Commission, 2021. Modernising Communications Offences: A final report. [accessed 28 July 2023].
[1489] 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].
[1490] See chapters 6E: Harassment, stalking, threats and abuse offences and chapter 6G: Controlling or coercive behaviour offence
[1491] Law Commission, 2021. Modernising Communications Offences: A final report. [accessed 26 July 2021].
[1492] Q21 (Table 1527). Source: Ofcom, 2023. Experiences of using online services. [. [accessed 1 August 2023].

Online dating services

6S.24 There is also evidence that cyberflashing is particularly prevalent on online 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.[1493] 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.[1494]

Risk factors: User base

User base size

6S.25 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 chapter 6W: Context to understand risk dynamics (found in the Annex of this document).

User base demographics

6S.26 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.

6S.27 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.

6S.28 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%).[1495]

6S.29 The evidence suggests that cyberflashing is most prevalent among young people; 47% of women aged 18-24 have been cyberflashed.[1496]

6S.30 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%).[1497]

6S.31 There is some evidence that cyberflashing is more prevalent among the **LGBT+ community**. 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.[1498]

[1493] 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].

[1494] McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. Bristol: Bristol University Press. [accessed 21 September 2023].

[1495] 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].

[1496] 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].

[1497] Ofcom, Online Experiences Tracker 2021-2022. Comprises Wave 1 and 2 combined data set.

[1498] Ofcom, Online Experiences Tracker 2021-2022. Comprises Wave 1 and 2 combined data set.

Risk factors: Functionalities and recommender systems

User identification

User profiles

6S.32 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 been able to create another account and an associated user profile to continue their offence.[1499]

Anonymous user profiles

6S.33 The ability to create anonymous user profiles could be a risk factor for cyberflashing offences. Research has found that anonymity creates a disinhibition effect[1500] which could lead perpetrators to engage in harmful behaviour such as cyberflashing.

User networking

User connections and user searching

6S.34 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.[1501]

6S.35 Victims and survivors describe unknown perpetrators making contact with them on social media in order to cyberflash.[1502] Several high-profile women have also shared their experiences of being cyberflashed by men who made contact with them on social media.[1503]

User communication

Direct messaging

6S.36 Direct messaging is a risk factor for cyberflashing offences. Victims and survivors describe perpetrators sending messages which contain unsolicited sexual images.[1504] Sixty-six per cent of women who have experienced cyberflashing and 59% of men were using a text or messaging app at the time.[1505]

[1499] 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].

[1500] 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].

[1501] McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. Bristol: Bristol University Press. [accessed 21 September 2023].

[1502] 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].

[1503] BBC, 2023. Emily Atack: Asking For It?, 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].

[1504] McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. Bristol: Bristol University Press. [accessed 21 September 2023]; Revealing Reality, 2022. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 28 July 2023].

[1505] 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].

Ephemeral messaging

6S.37 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-permanent images. Young people in particular report being cyberflashed via ephemeral messaging services.[1506]

Posting content (images and videos)

6S.38 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.[1507]

Livestreaming

6S.39 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.[1508]

Content storage and capture

Capturing images

6S.40 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

6S.41 No specific evidence was found on how business models may influence risks of harm to individuals for this offence.

[1506] 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. Not Just Flirting: The unequal experiences and consequences of nude image-sharing by young people. [accessed 28 July 2023].

[1507] 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].

[1508] 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].

Part 2: Search services

6T. Search services

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

Introduction

6T.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.

6T.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. To do so, they index most of the pages across the 'clear web'.[1509] [1510]

6T.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. This can happen unless mitigations are in place that specifically prevent illegal content from being returned in search results; if illegal content has been indexed, it can be encountered by users.

6T.4 The rationale for having a separate chapter on search services are the differing risk assessment duties that the Online Safety Act (the Act) places on Ofcom, and the fundamental differences between search services' limited functionality compared to user-to-user (U2U) services.

Service types

6T.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.

6T.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 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.[1511] 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).

[1509] Indexing is the process of collecting, parsing, and storing of data to facilitate fast and accurate information retrieval.

[1510] The 'clear web' refers to publicly accessible websites that are indexed by search engines.

[1511] A search service is defined in section 3 of the 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

6T.7 Services that allow users to search more than one website or database may be a 'search service'[1512]. 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.[1513] A provider of a search service is the entity that exercises control over the operations of the search service (for example, the way in which a user inputs a search query and the presentation of search results, or if they obtain their search results from other search providers).[1514]

6T.8 Ofcom has identified several types of search services based on the definitions in the Act and how search services operate.

a) **General search services**: General search services enable users to search the contents of the web by inputting search queries on any topic and returning results. There are two types of general search service:

i) **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 details of which are proprietary and not publicly known.[1515] There are a small number of large general search services that do their own crawling and indexing, which then provide search results to downstream general search providers. There are also smaller search services which do their own indexing.

ii) **Downstream general search services:** As a type of general search service, downstream general search services provide access to content from across the web, but they are distinct in that they obtain (or supplement) their search results from those general search services which rely solely on their own indexing. Depending on their contract with a general search service, downstream general search services may have limited control over how search services are displayed.[1516] Downstream general services often distinguish themselves from general search services by offering a social purpose (e.g. Ecosia), additional privacy (e.g. DuckDuckGo), or differentiated search features.

b) **Vertical search services:** Also known as 'speciality search engines', enable users to search for specific topics, or products or services offered by third party providers. They operate differently from general search services. Rather than crawling the web and indexing webpages, they present users with results only from selected websites with

[1512] Sections 3 and 229 of the Online Safety Act.

[1513] Refer to section 4 of the Online Safety Act and Schedule 1 to the Online Safety Act.

[1514] Section 226 of the Online Safety Act.

[1515] 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.

[1516] 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].

which they have a contract, and an API[1517] or equivalent technical means is used to return the relevant content to users. Common vertical search services include price comparison sites and job listing sites.

6T.9 There are also two ways in which content generated by GenAI models and/or GenAI service may fall within scope of the Act's search duties. 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[1518] Second, a conventional search service could use a GenAI model to augment its search engine and facilitate the delivery of its search results.[1519] 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

6T.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’).[1520]

6T.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[1521] that can occur to an individual as a result of:

a) a user directly encountering illegal content in or via the search results1522 of a search service; or
b) 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.

6T.12 Content is illegal if it amounts to a relevant offence. This includes both priority offences[1523] and other relevant offences,[1524] including communications offences.[1525]

[1517] Application Programming Interface is a way for two or more computer programs to communicate with each other.

[1518] For instance, a GenAI service could draw on more than one website or database by providing real-time information from plug-ins.

[1519] 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.

[1520] Section 98 of the Online Safety Act.

[1521] Section 234 of the Online Safety Act.

[1522] Section 57 of the Online Safety Act defines both search content and search results. Broadly, search content is content encountered in or via a search result (i.e.: 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 this definition.

[1523] Offences listed under Schedules 5 (’terrorism offences’), 6 (’child sexual exploitation and abuse offences), and 7 (priority offences) to the Act.

[1524] 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.

[1525] 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.

6T.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.

6T.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** (Volume 5, Chapter 26).

How harm manifests on search services

Risk of harm to individuals presented by illegal content on search services

6T.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 which 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.

6T.16 Although the mechanisms by which illegal content can manifest itself may be different on search services compared to U2U services, the impact on individuals is comparable. We therefore recommend readers to refer to the evidence summarised in Part 1: User-to-user services chapters to understand how the risks of harms manifest.

6T.17 While the evidence base surrounding GenAI and its ability to lead users to illegal content on search services is currently limited, there is some evidence which suggests that GenAI technologies on search services are at risk of amplifying the risk of harm to individuals.

6T.18 Research indicates that 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.[1526] There is also evidence illustrating how such services could be used to share malicious links and steer search results towards manipulated content.[1527]

Evidence of risk factors on search service

6T.19 Evidence reviewed for the Register was concerned with one key area: whether the characteristics[1528] – 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 liable to increase the risk of harm to individuals.

[1526] 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].

[1527] 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].

[1528] For further information on the characteristics, see chapter 6: Introduction to the Register.

6T.20 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.

Risk factors: Service types

6T.21 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:

a) General and downstream search services present users with access to webpages from across the entire clear web, while vertical search services have a far narrower scope, presenting content to users from pre-determined locations on the web.

b) Vertical search services are likely to have a lower level of risk of harm than general or downstream search services. This is because vertical search services typically only focus on a specific segment of online content and draw results from pre-determined websites that 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.[1529]

General and downstream search services

6T.22 The evidence suggests that using search services is an effective way for users to access the illegal content they are looking for. Although there is limited evidence on the volume of illegal content directly accessible via search services, the evidence covers a broad range of content that may amount to a relevant 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.

6T.23 There is evidence that search services can be used to access content that amounts to an offence related to terrorism.[1530] Studies also show that search services have returned content that relates to hate offences in relation to certain groups, notably antisemitic content.[1531]

6T.24 Evidence suggests that search services are used to access extreme pornography, with users searching for relevant terms that describe extreme pornographic content which would be

[1529] 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.

[1530] 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].

[1531] 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].

prohibited.[1532] There is also evidence that searches for extreme pornography allow users to be led to videos directly from the results page.[1533]

6T.25 Search engines are identified as one of the most common methods of finding CSAM (child sexual abuse material) online,[1534] 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]".*[1535]

6T.26 There is also evidence showing the ways in which search engines can be used to access websites hosting illegal items, such as drugs,[1536] firearms[1537] or articles used in fraud offences such as stolen credit card details.[1538]

6T.27 Search results have been shown to present users with content related to suicide and self-harm, including potentially illegal content. Studies on self-harm patients and male suicide victims found that both these groups use search engines to access pro-suicide content and to research methods.[1539]

6T.28 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. 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 by users to increase the likelihood of illegal content being displayed to users. For example, the tactic of keyword stuffing (filling a web page with keywords or numbers in an attempt to manipulate

[1532] 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+.

[1533] 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+.

[1534] 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].

[1535] 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].

[1536] Commission On Combating Synthetic Opioid Trafficking, 2022. Technical Appendixes. [accessed 17 October 2022].

[1537] 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].

[1538] 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].

[1539] 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].

rankings in search results)[1540] has been identified in research looking at how easily illegal content relating to fraud can be accessed via search services.[1541]

Risk factors: User base

6T.29 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

6T.30 The size of a search service's user base is a risk factor; all else equal, a service with a large user base could lead to more people being exposed to harmful content.

User base demographics

6T.31 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.

6T.32 As evidenced in 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.

Risk factors: Functionalities

6T.33 These include the underlying potential for illegal content on webpages indexed by 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).

6T.34 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

6T.35 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.

Image search

6T.36 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 search.[1542] While published evidence for other buying/selling offences is limited, it

[1540] Google, n.d. Spam policies for Google web search. [accessed 26 September 2023].

[1541] Ofcom, 2023. Articles and items for use in the commission of fraud – accessibility via search services. [accessed 26 September 2023].

[1542] RAND, 2022. Commission On Combating Synthetic Opioid Trafficking. [accessed 26 September 2023].

is possible that the reverse image search functionality also presents opportunities to access content relating to other prohibited items.

Search prediction and personalisation

6T.37 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 of offences, unless effective mitigations are in place to prevent this, or indexed content is blocked.

6T.38 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.[1543] Evidence has also found that search bar predictions have recommended hateful or racist search queries.[1544]

6T.39 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.[1545]

Risk factors: Business models and commercial profiles

Revenue models

6T.40 General search services typically generate revenue using an advertising-based model.[1546] 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.[1547] We understand that downstream general search services also earn revenue through advertising.[1548]

6T.41 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.

6T.42 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

[1543] For more detail see chapter 6D: Encouraging or assisting suicide (or attempted suicide) or serious self-harm offences. Google search results as of 9th March 2022, examples provided to Ofcom by the Samaritans.

[1544] 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.

[1545] Ofcom, 2023. Articles and items for use in the commission of fraud – accessibility via search services. [accessed 26 September 2023].

[1546] Some search engines use a subscription model in lieu of advertising to generate revenue, although this is exceedingly rare.

[1547] Competition and Markets Authority, 2020. Online platforms and digital advertising. [accessed 26 September 2023].

[1548] Competition and Markets Authority, 2020. Online platforms and digital advertising: Market study final report. [accessed 26 September 2023]; Australian Competition and Consumer Commission, 2021. Digital platform services inquiry. [accessed 26 September 2023].

coercive control.[1549] Although these devices are illegal, the research suggests that they are sometimes marketed as parental safeguarding tools.[1550]

6T.43 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.

6T.44 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.[1551] 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*".[1552]

Growth strategy

6T.45 Developing an expansive search index is a core component in a search service's growth strategy. A service's approach to expanding its index may influence the risk of harm. If services 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 service is looking to develop its own index rapidly in order to compete with more established services, prioritising this over user safety.

Commercial profile

6T.46 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.[1553]

6T.47 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 search results. For instance, they may have limited technical skills, financial resources or lack access to data from useful third parties (i.e, CSAM URL lists from the IWF).

6T.48 We recognise that these services often source their search results from larger providers (e.g. Google and Bing) due to their limited capacity/resources to develop an index independently.[1554] These 'downstream' services may have the benefit of importing their

[1549] 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].

[1550] 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].

[1551] Bradshaw, S, 2019, *Disinformation Optimised: Gaming Search Engine Algorithms to Amplify Junk News*, Internet Policy Review, 8(4). [accessed 26 September 2023].

[1552] Google Threat Analysis Group (Huntley, S.), 2022. TAG Bulletin: Q2 2022. [accessed 26 September 2023].

[1553] 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.

[1554] We understand that the level of control that a downstream general search service has over its search results depends on the contract the provider has with the service from which it buys the results. The nature of these contracts is not publicly known and is likely to differ from service to service. If a downstream service does not have full control over its search results, this may limit its ability to protect its users.

provider’s index safety profile, but may also lack the autonomy to remove harmful links themselves (as they have limited control over their search results).

Part 3: Governance, systems and processes

6U. Governance, systems and processes

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

Introduction

6U.1 As part of this risk assessment for illegal content, we have assessed the characteristics of a service relating to governance, systems and processes (GSP) separately. This is because the analysis of risk arising from these characteristics cuts across the different kinds of illegal harms. Governance, systems and processes are also important in offering ways for services to mitigate the risk of different kinds of illegal harms to users on their service.

6U.2 However, after reviewing the evidence, 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.

6U.3 This section provides services additional information of these risks. We have grouped them as per our Codes of Practice, which can be found in Volume 4 of this consultation. These are:

a) Governance
b) User access
c) Content moderation
d) Service design and user support
e) Reporting, complaints and appeals

6U.4 The Online Safety Act (the Act) does not provide a definition for '**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".*[1555]

6U.5 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".*[1556] For the purposes of this risk assessment, we interpret this to mean any series of actions taken by a service, including actions that mitigate the risks of harm arising from illegal content being encountered that may not have been addressed elsewhere in the Register.

[1555] Milliman, 2023*. Report on principles-based best practices for online safety Governance and Risk Management*. This report was commissioned by Ofcom and is published alongside this consultation. 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.

[1556] Section 236 of the Online Safety Act.

Evidence of risks of harm to individuals arising from governance, systems and processes

6U.6 Many governance arrangements and systems and processes that services put in place play important roles in mitigating the risk of harms to their users, as identified throughout this Register of Risks (Register).

6U.7 For example, in the U2U Hate and Terrorism chapters, we highlighted how the ability to share illegal content rapidly and widely on a U2U service can increase the risks of harm to individuals. One such instance is where users have the ability to livestream terrorist acts on a U2U service. Here, content moderation[1557] systems can play a crucial role for services to detect and remove this kind of content swiftly.

6U.8 Although GSPs of this kind can reduce risk, services should consider how they implement them, as this can also affect wider online safety decisions and responsibilities within a service. In fact, if implemented ineffectively, GSPs can have the inverse effect of increasing the risk to users in some contexts.

6U.9 The following sections provide services with an overview of these dynamics across our Codes of Practice areas. We expect services to use this information when considering how their own current or future safety measures will affect users' safety.

Organisational structure and governance

6U.10 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.

6U.11 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.

6U.12 Further information as to how services can mitigate the risks outlined here can be found in our Codes (Volume 4 of this consultation).

[1557] Content moderation refers to when a service reviews content to decide whether it is permitted on its service. It can be done by humans, automatically, or a combination of the two. It includes the identification and assessment of content and any actions taken on content. This can include rules imposed on content, the human labour and technologies required, and the institutional mechanisms of enforcement and appeal that support it. Roberts, T., 2017. Content Moderation, *Encyclopedia of Big Data*, pp.1-4. [accessed 21 September 2023]

Governance arrangements

6U.13 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 monitor the effectiveness of a company's risk and governance practices.[1558] 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 with a more risk-averse attitude towards investments, as non-executive directors may be more concerned about their reputations.[1559]

6U.14 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.[1560] 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.[1561] Analysis of a high-profile safety incident in the aerospace industry has in fact shown 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.[1562]

Senior accountability and responsibility

6U.15 Our analysis has 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.

6U.16 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.

6U.17 Evidence from other sectors indicates that inadequate leadership is one of the key contributors to poor safety outcomes. For example, there is evidence from financial services and health and safety sectors in high-profile cases.[1563] Senior leadership failures in financial

[1558] OECD, 2023. G20/OECD Principles of Corporate Governance. [accessed 6 October 2023]; Milliman, 2023. *Report on principles-based best practices for online safety Governance and Risk Management*. This report is published alongside this consultation.

[1559]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 4 October 2023].

[1560] 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 24 August 2023].

[1561] 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].

[1562] 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 05 May 2023].

[1563] 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].

services in relation to the 2008 financial crisis are a case in point that resulted in reduced oversight and excessive risk-taking.[1564]

Internal assurance and compliance functions

6U.18 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*." [1565] 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.[1566] For these reasons, internal assurance and compliance functions can be effective in ensuring that there is adequate oversight over risk management.

6U.19 Evidence from examples of high-profile organisational failures highlight the importance of effective internal controls in managing and mitigating a range of risks.[1567] Our analysis points to weak or absent controls as a key contributory factor to organisational failure, especially in major corporate scandals.

6U.20 We found evidence supporting the hypothesis that poor internal controls played a role in high-profile instances of organisational failures related to fraud,[1568] data integrity[1569] and product safety[1570] across other sectors.

[1564] 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].

[1565] European Commission, 2022. Public consultation on the strengthening of the quality of corporate reporting and its enforcement: summary report. (accessed 03 May 2023).

[1566] 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].

[1567] Di Miceli Da Silveira, A., 2011. 'Corporate Scandals of the Earlier 21st Century: Have We Learned the Lessons?' (accessed 03 May 2023).

[1568] 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.

[1569] 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].

[1570] *"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].

Staff incentives, policies and processes

6U.21 Our analysis suggests that there is an increased risk of harms to users if staff across the service are not adequately trained on compliance.[1571] 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.

6U.22 Without efforts to align safety objectives across different areas of a service, it is possible that staff will not understand how a service is approaching regulatory compliance, or how it manages and mitigates risks of illegal content being displayed to users. This is supported by evidence of how the absence of compliance training programmes has contributed to serious corporate scandals.[1572]

User access

6U.23 Some user-access-related functionalities may in some instances increase the risks of harm.[1573] This is outlined in detail in offence-specific chapters of the Register of Risks, such as chapter 6H: Drugs and other psychoactive substances, chapter 6C: Child sexual exploitation and abuse (CSEA) and chapter 6E: Harassment, stalking, threats and abuse/Threatening communications.

6U.24 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.

6U.25 Further information as to how services can mitigate the risks described here can be found in our Codes (Volume 4).

Account strikes and blocking procedures

6U.26 Evidence has demonstrated that risks of harm may increase when account strikes and blocking procedures are ineffective or unclear;[1574] for example, when services do not remove or limit violative user profiles such as those reported from users who have received abuse[1575] or user profiles supplying drugs.[1576] 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.[1577]

[1571] 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].

[1572]Primbs, M. and Wang, C., 2016. Notable Governance Failures: Enron, Siemens and Beyond. *Comparative Corporate Governance and Financial Regulation. 3.* [accessed 21 September 2023].

[1573] 'User access' refers to a user's entry into a service and ability to use the functionalities present on that service.

[1574] Thorn, 2021. Responding to online threats: minors' perspectives on disclosing, reporting and blocking. [accessed 21 September 2023].

[1575] 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].

[1576] 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 2022].

[1577] Business Insider, 2015. Transgender Tinder Users Reported and Banned. [accessed 4 October 2023].

6U.27 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[1578] [1579] and incitement to violence,[1580] but has also been observed as a trend for violative behaviour in general.[1581] 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 chapter 6B: Terrorism.

6U.28 While transparent systems and processes can increase user understanding of 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.

Content moderation

6U.29 Poor content moderation may in some instances facilitate illegal content being encountered and shared on both U2U and search services; this increases the risks of harm to individuals. Offence-specific risks of harm relevant to content moderation are outlined in different chapters in the Register; the most prominent examples are in chapter 6F: Hate offences, chapter 6B: Terrorism and chapter 6C: CSEA. One such example relates to the livestreaming of terrorist attacks or of child sexual abuse.

6U.30 Content moderation, whether automated, human, or a combination of both, is put in place by services to mitigate these risks. However, additional risks may arise if content moderation systems and processes are implemented poorly. We outline this in our analysis below.

6U.31 Further information as to how services can implement content moderation effectively and mitigate the risks described here can be found in the Codes of Practice (Volume 4).

Resourcing and time constraints

6U.32 As mentioned above, our analysis of the evidence on both U2U and search services suggests that content moderation systems which are poorly designed, deployed and resourced may increase the risk of illegal content appearing on services.

[1578] 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.** Source: Meta, 2021. Understanding the intentions of Child Sexual Abuse Material (CSAM) sharers. [accessed 27 March 2023].

[1579] 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].

[1580] 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].

[1581] In a recent 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].

6U.33 Firstly, services with less stringent moderation may be viewed as preferred spaces for discussions about terrorism and hate, when compared to services with more moderation.[1582] This is part of an observed trend, where services with less stringent moderation, and with fewer staff, have, on average, had higher volumes of terrorism content than services with stringent moderation.[1583]

6U.34 Secondly, research has shown that the reduction of staff for content moderation in a large service led to a major increase in the quantity of antisemitic content on the service.[1584] [1585]

6U.35 It has also been found that the minimal moderation of a messaging service, allowing users to create groups and channels, had "lowered the hurdle for engaging 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.[1586]

6U.36 Time pressures[1587] on human moderators and the poor resourcing of moderation[1588] can also increase the risk of human error in moderation decisions. Similarly, inadequate training or quality assurance, 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.[1589] In response to our Call for Evidence, Refuge stated that "*to the untrained eye, tech abuse can often be hard to recognise without an understanding of the broader context of domestic abuse and coercive control*".[1590]

[1582] The Institute for Strategic Dialogue (O'Connor, C.), 2021. Gaming and Extremism: The Extreme Right on Twitch. [accessed 4 October 2023].

[1583] Tech Against Terrorism, 2023. TCAP Insights. Patterns of online terrorist exploitation. [accessed 4 October 2023].

[1584] CASM Technology and ISD, 2023. Antisemitism on Twitter Before and After Elon Musk's Acquisition. [accessed 4 October 2023].

[1585] 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].

[1586] HOPE not hate and the Antisemitism Policy Trust, 2021. Antisemitism and Misogyny: Overlap and Interplay. [accessed 4 October 2023].

[1587] A report by Demos highlighted that human content moderators have to make decisions in minutes, often about content in a language or from a context they do not understand, making mistakes inevitable. Source: Demos (Krasodomski-Jones, A.), 2020. Everything in Moderation: Platforms, communities and users in a healthy online environment.[accessed 4 October 2023].

[1588] 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."* Similarly, in late 2022, the Anti-Defamation League (ADL), noted an increase in antisemitic content on the same service and a decrease in the moderation of antisemitic posts. Source: ADL, 2022. Extremists, Far Right Figures Exploit Recent Changes to Twitter. [accessed 3 August 2023]; CASM Technology and ISD, 2023. Antisemitism on Twitter Before and After Elon Musk's Acquisition. [accessed 3 August 2023].

[1589] 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 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.

[1590] Refuge response to Ofcom's 2022 Call for Evidence.

6U.37 However, it is not only ineffective content moderation that can lead to risk. Content moderation can be a challenge simply due to the large volumes of content that need to be sifted through; a point a large search service acknowledges*: "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"*.[1591]

Service design

6U.38 Service design[1592] 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 or search services. Offence-specific risks of harm associated with service design are outlined in different chapters of this Register, and the most prominent examples are in chapter 6D: Encouraging or assisting suicide or serious self-harm and chapter 6L: Extreme pornography. Such examples relate to how vulnerable users may be recommended content that is increasingly harmful and potentially illegal. Similarly, users may be led to illegal content within a few clicks from their query on a search service (for further information, see chapter 6T on risks of harm to individuals on search services).

6U.39 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 chapter 6O: Fraud and financial services and chapter 6P: Foreign interference of this Register. 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 chapter 6B: Terrorism.

6U.40 However, services use design systems and processes to mitigate the risks mentioned above; if these are implemented poorly, the risks of harm may increase. We outline this in our analysis of 'recommender systems', 'verification and labelling of user accounts', 'predictive search' and 'web indexing' below.

6U.41 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 4).

[1591] Google, 2020. Information quality & content moderation. [accessed 21 September 2023].

[1592] 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.

Recommender systems

6U.42 There are a number of ways by which illegal content can be disseminated on an online service; content recommender systems[1593] 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.

6U.43 Evidence has shown that 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 services making design changes 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. [1594] While the focus of these studies tends to be on harmful content, our view is that illegal content would be disseminated in a similar way, given how recommender systems rank and curate content.[1595]

6U.44 To back up these claims, there is specific evidence of recommender systems disseminating CSAM and terrorism. 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.[1596] Evidence also showed how the recommender system of one social media service had recommended Islamic State accounts to users.[1597] 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.[1598]

Verification and labelling of user accounts

6U.45 Many online services run schemes to verify the accounts of notable users (such as celebrities, political figures, state media or financial institutions) and say the label can help other users identify the 'authentic' presence or profile of these users on the service.[1599] The lack of these measures may therefore result in 'fake user profiles', which are detailed in the

[1593] Content recommender systems are 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.

[1594] Global Internet Forum to Counter Terrorism, 2021. Content-Sharing Algorithms, Processes, and Positive Interventions Working Group: Part 1. [accessed 27 September 2023]

[1595] Ofcom, 2023. Evaluating recommender systems in relation to illegal and harmful content. [accessed 4 October 2023].

[1596] Cook, J. and Murdock, S., 2020. YouTube Is A Pedophile's Paradise, *HuffPost, 20 March*. [accessed 4 October 2023].

[1597] 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].

[1598] 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. 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].

[1599] 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.

Register. Recently, some online services have begun to offer monetised schemes whereby users can be verified, have their account labelled and get access to special features in return for payment.

6U.46 However, confusion about user labelling and verification schemes, 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 verified 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.[1600] It has also been reported that verification schemes may be misused to impersonate conspiracy theorists and far-right activists.[1601] [1602] [1603] [1604]

Search

Predictive search functionalities

6U.47 Predictive search functionalities may inadvertently lead a user to illegal content. Respondents to our 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[1605], or to hateful and racist content that can lead eventually to illegal content.[1606] One respondent to our Call for Evidence stated that "*search services have been 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*."[1607] The latter is also substantiated by an investigation by The

[1600] Farrell, L, 2023. Martin Lewis issues warning over fake Twitter account after major change to app. *Daily Record,* 4 April. [accessed 4 October 2023].

[1601] Sardarizadeh, S, 2022. Twitter chaos after wave of blue tick impersonations. BBC News, 12 November. [accessed 4 October 2023].

[1602] 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].

[1603] 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].

[1604] 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].

[1605] Samaritans' response to Ofcom's 2022 Call for Evidence. Samaritans is the UK and Ireland's largest suicide prevention charity.

[1606] 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*".

[1607] Antisemitism Policy Trust response to Ofcom's 2022 Call for Evidence.

Observer in 2016.[1608] There is also research that points to similar risks regarding child sexual abuse material (CSAM).[1609] [1610]

Web indexing

6U.48 Web indexing may also present risks of harm to individuals. Through the indexing process, search services act as a gateway to the entire contents of the web, including URLs and images that contain illegal content. Evidence indicates that this risk is heightened for CSAM, especially as search engines are a common way of finding this type of illegal content.[1611] [1612] Research has found that this material could be reached within three clicks.[1613]

6U.49 Research completed by Ofcom has found that content offering to supply articles for use in frauds is easy to find and prevalent in or via search results on some search services. Search queries used in this research returned large volumes of content within the first 20 search results, which we categorised as 'likely to be prohibited'.[1614] This research also generated a range of insights around the prevalence of specific terms associated with the sale of stolen credentials online. Notably, the research found that slang, coded language and more detailed search keywords or queries led to a higher proportion of content likely to be prohibited.[1615] Another important finding of the research is that search services can also direct users to such content on user-to-user services. Research commissioned by Cifas aligns with our own research, highlighting the prevalence of this type of content online and the use of specific terms associated with articles for use in frauds.[1616]

6U.50 Further information as to how services can implement service design effectively on search services, and mitigate the risks described here, can be found in the Codes of Practice (Volume 4).

[1608] 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].

[1609] 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].

[1610] 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].

[1611] 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].

[1612] 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].

[1613] 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].

[1614] Search terms used in this research returned large volumes of content within the first 20 search results, content which is likely to be examples of this priority offence. Ofcom, 2023. Online Content for use in the commission of fraud -accessibility via search services. [accessed 18 September 2023].

[1615] Ofcom, 2023. Online Content for use in the commission of fraud -accessibility via search services. [accessed 18 September 2023].

[1616] Cifas and Forensic Pathways, 2018. Wolves of the Internet: Where do fraudsters hunt for data online? [accessed 25 September 2023].

Reporting, complaints and appeals

6U.51 Based on the analysis above and on the analysis in the Register, we see how illegal content can be encountered and shared in a number of ways on U2U and search services, thereby increasing the risk of harms to individuals.

6U.52 Our analysis has shown that services may put in place reporting and/or complaints mechanisms, plus an appeals process, to mitigate the risks of these harms, and to identify illegal content that has been undetected by content moderation systems. However, if these systems and processes are implemented poorly, they may themselves increase the risks of harm to individuals. For example, users may find it difficult to report if the process is too complicated, lengthy or unclear.

6U.53 Further information as to how services can implement reporting, complaints and appeals effectively, and mitigate the risks described here, can be found in the Codes of Practice (Volume 4).

Accessibility of reporting/complaints

6U.54 Our analysis of the evidence suggests that children in particular[1617] do not report or complain about violative content if the reporting channels are unclear. This can hinder content takedown, for instance, in the case of CSEA.[1618] In some cases, ineffective reporting and complaints systems result in services being unable to collect reliable information on the type of harmful content and behaviour appearing on their service, and therefore unable to take proactive action against further harm.[1619] There is a greater risk of illegal content remaining on the service for longer if users struggle to report/complain, increasing the likelihood of harm to more users who may then come across it or be subject to repeated exposure.

[1617] *"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.

[1618] 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].

[1619] 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].

6U.55 Ofcom's Video-Sharing Platforms Call for Evidence highlighted that services do operate dedicated reporting, such as the 'trusted flagger' programmes that enable law enforcers, civil society, charities and other stakeholders to alert services to harmful content;[1620] but these are often applied inconsistently and may fail to address complex kinds of illegal harms like fraud.

Speed of action

6U.56 Evidence has shown that where users perceive a lack of action on a report, or an absence of any action at all,[1621] they are less likely to report in the future. This further increases the risk of harm to individuals.[1622] 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 trauma they may feel.[1623] A study with LGBT+ users found that they fear they will not be taken seriously when reporting content.[1624]

6U.57 On the other hand, benchmarking quick turnarounds for reports arbitrarily may prevent appropriate assessment of illegal content. There is a risk that if services decide to set themselves arbitrary deadlines to address complaints, they may not assess them rigorously enough. This has been highlighted as a concern in EU proposals for content removal of terrorism content within specific timeframes.[1625]

[1620] Tech UK response to 2020 Video-Sharing Platform Regulation Call For Evidence , p.3. [accessed 31 August 2023]

[1621] Ofcom, 2022. Just one in six young people flag harmful content online. [accessed 25 September 2023].

[1622] 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].

[1623] 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].

[1624] Galop (Hubbard, L.), 2020. Online Hate Crime Report 2020: Challenging online homophobia, biphobia and transphobia. [accessed 4 October 2023].

[1625] 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].

6V. Annex: 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.[1626]
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.[1627]
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.[1628]
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.
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.

[1626] Section 98(11) of the OS Act.
[1627] Section 207(1) of the OS Act.
[1628] Section 201 of the OS Act.

Term	Definition
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.[1629]
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 where a user's identity[1630] is unknown to other users, for example 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.[1631]
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.

[1629] Section 57(3) of the OS Act.

[1630] Identity refers to an individual's formal or officially recognised identity.

[1631] 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.

Term	Definition
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.
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.
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.

Term	Definition
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.[1632] 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).[1633] 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.[1634]
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.

[1632] 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.

[1633] Section 233(3) of the OS Act.

[1634] 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.

Term	Definition
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,[1635] 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.
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.[1636]
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.
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.[1637]
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[1638] channels.

[1635] See 'advertising-based revenue model' in business models for more information.

[1636] This for instance includes 'liking' or 'disliking' a post.

[1637] 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.

[1638] See content audiences for definition of open and closed channels of communication.

Term	Definition
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.
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 represents a collection of identifying information about a user conveyed to other users of the service. This can include information that may be displayed to other users such a images, usernames, and biographies.[1639, 1640]
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.

[1639] User profiles are distinct from user accounts, which are representations of a user in a service's' information system. They may contain information required for registration to a particular service that are often attributes of a user's identity such as name, age, contact details and preferences.

[1640] 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.

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. [1641]
Commercial profile	Size of the service in terms of capacity,[1642] 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.
High-capacity services	Services with a large number of employees and/or revenue[1643]
Low-capacity services	Services with a small number of employees and/or revenue[1644]
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.[1645]
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.[1646]
User base demographics	Demographic make-up of the user base, including selected characteristics, intersectional dynamics and other relevant demographic factors.

[1641] '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.

[1642] In terms of number of employees and/or revenue.

[1643] 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.

[1644] 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.

[1645] Section 4 of the Equality Act 2010.

[1646] 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.'

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. 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.

Term	Definition
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.
Adult services	User-to-user service type describing services that are primarily used for the dissemination of user-generated adult content.
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.[1647] 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.

[1647] Certain service types have been selected because our evidence suggests that they can be used to facilitate or commit relevant offences.

Other terms

Term	Definition
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.
Blockchain	A decentralised, distributed ledger that stores the record of ownership of digital assets.[1648]
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.[1649]
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.

[1648] Builtin (Daley, S.), 2022. What Is Blockchain? [accessed 18 June 2023].

[1649] Ministry of Justice, 2022. New laws to better protect victims from abuse of intimate images. [accessed 3 August 2023].

Term	Definition
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.
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.[1650]
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.

[1650] Metropolitan Police, n.d. Sextortion. [accessed 4 August 2023].

6W. Annex: Context to understand risk factor dynamics

6W.1 The chapters in this Register looked at the risks of harm to individuals for each kind of illegal harm. We were also able to draw out broader dynamics of risk, and explain the wider context of perpetrator behaviour online, the effect of new emerging technologies on kinds of illegal harm, and the importance of media literacy in preventing, or exacerbating if a perpetrator has high media literacy levels, some harm to individuals. We summarise some key points below.

U2U services

A range of services can attract perpetrators

6W.2 Our provisional 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 be a risk factor for 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.

6W.3 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.

6W.4 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. [1651] 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 Consultation Annex 5: Service Risk Assessment Guidance).

[1651] 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.

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

6W.5 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 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.

6W.6 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 chapter 6B: Terrorism offences for further information)

A service's commercial and revenue models, as well as growth strategy can increase risk

6W.7 A service's revenue model may be relevant when considering risks of harm. They may have features that can be misused by perpetrators e.g. ad targeting may be misused by perpetrators to reach potential victims.[1652] In addition, revenue models may create financial incentives that – intentionally or unintentionally – lead to business decisions that promote or tolerate illegal activity. For example, the way in which an advertising revenue model seeks to increase user engagement may increase exposure to illegal content if the content improves this and, in turn, attract adverting revenue.

6W.8 At the same time, commercial pressure from revenue sources or the wider ecosystem (e.g. payment providers or investors)[1653] can generate incentives to clamp down on illegal content. Therefore, the link between advertising-funded business models and the risk of harm is not straightforward. The same applies for other types of revenue models.[1654]

6W.9 Businesses with different commercial profiles can have weak risk management, which makes them targets for perpetrators. For example, services that are low-capacity, at an early-stage or have a fast-growing user base may face an increased risk of harm. They may be targeted by perpetrators due to the assumed limitations of their risk management processes. In particular:

[1652] 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 as part 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.

[1653] 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 is published alongside the consultation.

[1654] We use advertising revenue model as an example to illustrate how revenue models may be relevant when considering risk, however, the same logic applies to other revenue models such as subscription, transaction fees, donations, etc. While there is more evidence in the Register on risks related to advertising revenue models, we do not rule out risks under other revenue models; there is insufficient evidence to conclude that advertising revenue models pose unique risks. In fact, some of the risks identified in the register under advertising revenue models can be applicable to other revenue models.

a) Low capacity and early-stage services are likely to have limited technical skills and financial resources to introduce effective risk management, compared to more mainstream services. In addition, they are likely to seek growth, which may affect their incentives to have effective risk management in place.

b) A fast-growing user base may negatively affect a service's ability to moderate effectively, given the increased scale and sophistication of the moderation technologies and processes required to keep track of a fast-growing user base (particularly since the sources of risk, and kinds of harms on the service, can change quickly as the user base develops).

6W.10 A service's growth strategy may also create tension with user safety. Services prioritising revenue growth can pose a higher risk of harm if they are incentivised to allocate their limited financial resources to support growth rather than fund moderation. In addition, if the growth strategy focuses on increasing the user base, the service may have a disincentive to moderate illegal content if it might attract a large number of new users quickly.

6W.11 While the evidence in the register on commercial profile and growth strategy is limited to very few kinds of illegal harm, we consider that the same logic applies across all kinds of illegal harms covered in the register.

Perpetrator behaviours

6W.12 Perpetrators can be wide-ranging which affects the risk of harm to individuals. 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 offender network. 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.

6W.13 Services should be mindful that perpetrators also service-hop, actively moving users to different services to commit or facilitate offences. Perpetrators may meet their victims and survivors, or other perpetrators, on one service and then move their interactions to another service. This move to another service may be driven by a service's characteristics; for example, in grooming, a perpetrator may choose to identify a child and initiate contact with them on a social media service, and then move their communication to a private 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's risk assessment.

6W.14 Service characteristics can be used as tools that help or enable a perpetrator to commit an offence. For example, user profiles and the information that is often displayed on them is used by perpetrators of grooming to identify vulnerable children to target. Although looking at user profiles is not illegal, in this case this characteristic enables a perpetrator to begin grooming.

6W.15 The channel of communication used to share content can lead to increased risks of harm to users. Perpetrators use both open and closed channels to cause harm to users. However, how and why they use them vary depending on whether the channel is open or closed. Open channels 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.[1655] Conversely, perpetrators are more likely to use closed channels for planning offences covertly or conducting illegal behaviour which they do not wish to be visible.[1656]

Service features and technologies

Any functionalities which facilitate communication at scale create both opportunities and risks

6W.16 Functionalities in general are not inherently positive nor negative. They facilitate communication at scale and reduce frictions 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. In contrast, functionalities can also enable the sharing of illegal material such as livestreams of terrorist atrocities or messages sent with the intent of grooming children.

6W.17 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.

The impact of emerging technologies - Generative Artificial Intelligence (GenAI) case study

6W.18 The effects of emerging technologies on the risks of harm to individuals will be difficult to measure due to the limited evidence available today. A service will need to consider how emerging technologies may affect perpetrator behaviour. Despite this, there is some evidence which suggests that GenAI technologies can be used to facilitate or commit harm on U2U services.

6W.19 Some studies highlight the use of GenAI to create 'deepfakes',[1657] which are subsequently used in foreign interference campaigns.[1658] This can include creating false or misleading videos of state figures. Our evidence also indicates that bots are used in foreign influence operations to produce or amplify content, including disinformation, under the direction of a person[1659] (see chapter 6P: Foreign interference offence and chapter 6Q: False communication offence). It is unclear whether these bots employ GenAI technologies. However, we believe that GenAI bots could be used in a similar manner.

[1655] This could be the case for fraud, hateful content or promotion of suicide content, for example.

[1656] 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.

[1657] Deepfakes refer to audio and visual content that has been manipulated to change how a person, object or environment is presented.

[1658] Brookings (Byman, D.L., Gao, C., Meserole, C. and Subrahmanian, V.S., 2023. Deepfakes and international conflict. [accessed 5 October 2023].

[1659] Bradshaw, S., Bailey, H. and Howard, P.N., 2021. Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation. [accessed 9 October 2023].

6W.20 GenAI may also be used to produce false communications[1660] and child sexual abuse material through the creation of deepfakes.[1661] For instance, the BBC has found that GenAI software is used to create *"life-like child sexual abuse material"*, which is then being shared through a different service that provides paid-subscriptions to user-generated content.[1662]

6W.21 Recent reports suggest that GenAI chatbots[1663] can offer advice that promotes eating disorders[1664] and encourages self-harm.[1665] 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[1666] which could be used to cause harm.[1667]

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

6W.22 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 enjoy and should be considered in the context of a service's ecosystem. However, they can also increase risk.

6W.23 Our evidence has found that two types of recommender systems could potentially increase risk: content recommender systems and network recommender systems.

a) Content recommenders may amplify risks of harm by (a) influencing the user journey: This involves pushing users into 'filter bubbles' and 'rabbit holes' or b) gaming of recommender systems by potential perpetrators; this is when bad actors abuse the algorithm by continuously posting and/or pushing extreme content.

b) Network recommenders may amplify risks of harm by helping offenders to find each other by connecting like-minded individuals and/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.

[1660] See chapter 6Q: False communications offence.

[1661] See chapter 6C: CSEA offences.

[1662] Crawford, A. and Smith, T., 2023. Illegal trade in AI child sex abuse images exposed, *BBC News,* 28 June. [accessed 5 October 2023].

[1663] An AI chatbot is an automated software program that uses artificial intelligence and natural language processing to simulate a conversation.

[1664] 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].

[1665] 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].

[1666] 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].

[1667] 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.

Users

Media literacy's influence on risks of harm

6W.24 Media literacy[1668], the ability to use, understand and create media and communications in a variety of ways[1669] can both contribute and limit the risk of harm.

6W.25 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.

6W.26 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.

6W.27 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.

[1668] Ofcom is mandated to promote media literacy. Section 11 of the Communications Act 2003 [accessed 28 June 2023].

[1669] Ofcom, 2023. Making Sense of Media Homepage. [accessed 28 June 2023].