Protecting people from illegal harms online

Annexes 12-16

Consultation

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

Contents

Annex

A12. Legal Framework Overview (Part A)

Introduction

A12.1 This Annex is in two sections. This first section sets out parts of the legal and regulatory framework under the Online Safety Act 2023 ('the Act') that are relevant to this consultation. It is intended to provide a high-level summary as context for our consultation proposals but is not a comprehensive outline of services' obligations under the Act. It focuses on the duties the Act places on Ofcom and online services in related to illegal content in particular, and we have not referred to aspects of the legal and regulatory framework which relate to the protection of children, which will be covered in our later consultation due to be published in 2024. You can find the full text of the Act here.[1]

A12.2 The second section of this Annex sets out Ofcom's and providers' duties relating to illegal content in detail.

A12.3 The Act places a number of duties on Ofcom and the online services who fall within scope of the new regime, namely user-to-user, search and pornography services. This consultation focuses on the first two categories of services.

The Online Safety Act 2023

Overview of the Act

Scope

A12.4 The Act provides for a new regulatory framework which has the general purpose of making the use of regulated internet services safer for individuals in the UK. To achieve this, the Act imposes duties which require providers to identify, mitigate and manage the risks of harm from illegal content and activity and content and activity that is harmful to children, as well as confers new functions and powers on Ofcom. Duties imposed on providers seek to secure, among other things, that regulated services are safe by design.[2]

A12.5 Internet services within scope of the new regulatory regime can be broadly grouped as:

a) a "user-to-user service", which means an internet service through which content that is generated, uploaded or shared by users may be encountered by other users of the service;[3]
b) a "search service", which means an internet service that is, or includes, a search engine;[4] or
c) a provider of internet services on which "provider pornographic content" is published or displayed.[5]

[1] https://www.legislation.gov.uk/
[2] The Act, section 1.
[3] The Act, section 3(1).
[4] The Act, section 3(4).
[5] The Act, section 79.

A12.6 Such services will only be in scope if they have "links to the United Kingdom"[6] and do not fall within Schedule 1 (exempt services). Regulated services have links to the UK if the service has a significant number of UK users or if UK users form one of the target markets or the only target market.[7] A service will also be considered to have links to the UK if it is capable of being used in the UK by individuals, and there are reasonable grounds to believe that there is a material risk of significant harm to individuals in the UK presented by user-generated content present on the service or search content of the service.[8]

A12.7 The Act establishes categories of regulated user-to-user and search services. Category 1 and Category 2B relate to different kinds of regulated user-to-user services, with Category 1 being the largest services. Category 2A relates to search services. The Secretary of State is responsible for setting threshold conditions for these categories based on the number of users of the service, its functionalities and other relevant factors.[9] Once the threshold conditions have been set, Ofcom is required to maintain and publish a register of the services in each category.[10]

A12.8 Please see Chapter 3 for further discussion of these provisions.

Provider duties

A12.9 The Act places duties of care (Part 3) and other duties (Part 4) on all providers of user-to-user services and search services. These include requirements to:

a) for user-to-user and search services:

i) carry out a suitable and sufficient illegal content risk assessment;[11]

ii) put in place systems and processes which allow users and affected persons to easily report illegal content and content that is harmful to children to the service provider;[12]

iii) operate a transparent and easy to use and access complaints procedure which allows for complaints of specified types to be made, including about illegal content;[13]

iv) have particular regard to the importance of protecting users' right to freedom of expression within the law, and protecting users from a breach of privacy, when deciding on and implementing safety measures and policies;[14]

v) put in place systems and processes designed to ensure that detected and unreported CSEA content is reported to the NCA.[15]

b) for user-to-user services only:[16]

i) take or use proportionate measures to prevent individuals from encountering priority illegal content; effectively mitigate and manage the risk of the service being used for the commission or facilitation of a priority offence; and effectively mitigate

[6] The Act, section 4(2)(a).
[7] The Act, section 4(5).
[8] The Act, section 4(6).
[9] The Act, Schedule 11.
[10] The Act, section 95.
[11] The Act, sections 9 and 26.
[12] The Act, sections 20 and 31.
[13] The Act, sections 21 and 32.
[14] The Act, sections 22 and 33.
[15] The Act, section 66.
[16] The Act, section 10.

and manage the risks of harm to individuals as identified in a service's most recent illegal content risk assessment; and

ii) use proportionate systems and processes designed to minimise the length of time for which any priority illegal content is present and to swiftly take down illegal content when the provider becomes aware of it;

iii) explain in clear and accessible terms of service how the service is protecting its users from illegal content and apply these terms of service consistently;

c) for search services only:[17]

i) take or use proportionate measures to effectively mitigate and manage the risks of harm to individuals as identified in a service's most recent illegal content risk assessment;

ii) use proportionate systems and processes designed to minimise the risk of individuals encountering priority illegal content and other illegal content that the provider knows about;

iii) explain in clear and accessible provisions in a public statement how individuals are to be protected from search content that is illegal content, and apply those provisions consistently;

d) for user-to-user services and search services likely to be accessed by children:

i) carry out a suitable and sufficient children's risk assessment in accordance with Schedule 3 and keep it up to date;[18]

ii) take or use proportionate measures to effectively mitigate and manage the risks of harm to children in different age groups as identified in a service's most recent children's risk assessment, and mitigate the impact of harm to children in different age groups presented by content that is harmful to children;[19]

e) in addition, user-to-user services likely to be accessed by children must notify Ofcom where a children's risk assessment identifies the presence of non-designated content that is harmful to children;[20] and operate the service using proportionate systems and processes designed to prevent children of any age from encountering primary priority content that is harmful to children and protect children in age groups judged at risk of harm from encountering other content that is harmful to children;[21]

f) search services likely to be accessed by children must also use proportionate systems and processes designed to minimise the risk of children of any age encountering primary priority content that is harmful to children, and minimise the risk of children in age groups judged to be at risk of harm from other content that is harmful to children.[22]

A12.10 In relation to illegal content, the Act defines this as "content that amounts to a relevant offence".[23] A relevant offence refers to a priority offence (terrorism offences,[24] offences related to child sexual exploitation or abuse[25] or other priority offences as specified in

[17] The Act, section 27.
[18] The Act, sections 11 and 28.
[19] The Act, sections 12(2) and 29(2).
[20] The Act, section 11(5).
[21] The Act, section 12(3).
[22] The Act, section 29(3).
[23] Section 59(2).
[24] The Act, Schedule 5.
[25] The Act, Schedule 6.

Schedule 7) or any other type of offence where the victim is an individual or individuals,[26] subject to certain exceptions.[27]

A12.11 Please see Chapter 2 for further discussion of these duties.

A12.12 Category 1 services are also subject to a number of additional duties, including requirements to abide by their terms of service and apply them consistently.[28] The Act also places specific duties on services in relation to certain pornographic content[29] and fees.[30] These duties are not yet engaged for the purpose of this consultation.

Ofcom's Codes of Practice and guidance

A12.13 Ofcom must issue Codes of Practice for regulated user to user and search services containing measures recommended for the purposes of compliance with certain duties referred to above, including the illegal content safety duties in sections 10 and 27.[31] In preparing these Codes of Practice, Ofcom must consider the principles and objectives set out Schedule 4 to the Act. Please see Chapter 24 for further discussion of these requirements.

A12.14 Where a Code of Practice exists, a provider of a regulated user-to-user service is to be treated as complying with a relevant duty if the provider takes or uses the measures described in the Code of Practice which are recommended for the purpose of complying with that duty (this is sometimes referred to as a "safe harbour"). In addition, providers are treated as complying with the cross-cutting duties regarding freedom of expression and privacy set out in sections 22 and 33 if they take or use such of the relevant recommended measures as incorporate safeguards to protect users' rights to freedom of expression and privacy. Providers may choose to take alternative measures to comply with the relevant duties rather than following the recommended measures in Codes.[32]

A12.15 Ofcom is further required to issue Illegal Content Judgements Guidance.[33]

Ofcom's duties relating to risk

A12.16 Ofcom must carry out risk assessments to identify and assess the risks of harm to individuals in the UK presented by:

a) illegal content on user-to-user services and the use of such services for the commission or facilitation of priority offences; and
b) illegal content that is search content encountered on search services.[34]

A12.17 Ofcom must also carry out a risk assessment to identify and assess the risk of harm presented by user-to-user and search services to children in the UK, in different age groups, by content that is harmful to children.[35]

[26] The Act, section 59(5).
[27] The Act, section 59(6).
[28] The Act, sections 71 and 72.
[29] Part 5 of the Act.
[30] Part 6 of the Act.
[31] The Act, section 41.
[32] The Act, section 49.
[33] The Act, section 193.
[34] The Act, section 98(1).
[35] The Act, section 98(1).

A12.18 We must also prepare and publish a Register of Risks that reflects the findings of our risk assessments[36] as well as Risk Profiles for user-to-user services and search services that relate to each risk of harm.[37]

A12.19 Ofcom is also required to issue guidance relating to how providers can comply with the risk assessment duties.[38]

A12.20 Please see Chapter 6 for further discussion of these duties.

Information gathering and enforcement

A12.21 The Act gives Ofcom broad powers regarding information gathering for the purposes of discharging our functions, including powers:

a) to require information generally from providers (by notice) for the purposes of exercising, or deciding whether to exercise, functions;[39]
b) to appoint a skilled person to provide a report to Ofcom for certain purposes relating to compliance;[40]
c) to require certain individuals to attend interviews and answer questions;[41] and
d) of entry, inspection and audit.[42]

A12.22 Ofcom is responsible for enforcing compliance with the duties in the Act on providers of regulated services. The duties on providers are generally enforceable by Ofcom where there are reasonable grounds for believing that a provider has failed, or is failing, to comply.[43]

A12.23 Sanctions for non-compliance may include requiring payment of a financial penalty of up to £18m or 10% of qualifying worldwide revenue.[44] In certain circumstances, Ofcom may apply to a court to take business disruption measures against platforms.[45] Ofcom may also, if certain conditions are met, issue notices requiring providers to use accredited technology or to develop or source technology to prevent users from encountering, or to identify and take down, terrorism content that is communicated publicly or CSEA content that is communicated publicly or privately.[46]

A12.24 Part 8 of the Act relates to appeals against Ofcom's decisions about the register under section 95 of the Act (regarding categorisation of services) and against Ofcom notices, while Part 9 relates to the Secretary of State's functions in respect of regulated services.

A12.25 Part 10 of the Act creates various new communications offences, some of which we refer to in our regulatory outputs. These offences include false and threatening communications offences and the offence of sending photographs or films of genitals.

A12.26 Further detail on the duties in the Act that are relevant to our consultation proposals is contained in the second section of this Annex.

---

[36] The Act, section 98(4).
[37] The Act, section 98(5).
[38] The Act, section 99.
[39] The Act, section 100.
[40] The Act, section 104.
[41] The Act, section 106.
[42] The Act, section 107.
[43] The Act, section 130.
[44] The Act, Schedule 13.
[45] See sections 144 to 148 of the Act.
[46] The Act, section 121.

The process for making Codes of Practice and guidance

Codes of Practice

A12.27 The Act specifies the procedure which applies to Ofcom when issuing, or amending, Codes of Practice.

A12.28 In the course of preparing a draft Code of Practice, Ofcom must consult various persons specified in section 41(6) and 41(7) of the Act. These include the Secretary of State; persons who represent services and their users; persons who represent the interests of children and those who have suffered harm as a result of matters to which the Codes relate; persons with expertise in equality issues, human rights, public health, criminal law enforcement, national security, innovation and emerging technology; and other public bodies such as the Information Commissioner and the Children's Commissioner, Domestic Abuse Commissioner and Commissioner for Victims and Witnesses.

A12.29 Once Ofcom has prepared a draft Code (or draft amendments to a Code), we must submit it to the Secretary of State.[47] The Secretary of State must either issue a direction under section 44 of the Act or lay the draft before Parliament. If either House of Parliament resolves not to approve the draft Code within the 40-day period,[48] Ofcom cannot issue that draft Code and must prepare another draft. If no such resolution is made, Ofcom must issue the draft Code in that form and it will come into force 21 days later.[49]

A12.30 The Secretary of State may direct Ofcom to modify a draft Code for exceptional reasons relating to national security, public health or safety or foreign relations or, in the case of a terrorism or CSEA Code, for reasons of national security, public health or safety or exceptional reasons relating to foreign relations.[50] If a draft terrorism or CSEA Code has been the subject of a review under section 47(2), or Ofcom has submitted a statement to the Secretary of State under section 47(3)(b) in respect of such a Code, the Secretary of State can only issue a direction to modify the draft for reasons of national security or public safety. A direction given under section 44 cannot require Ofcom to include any particular measure in a Code and must set out the Secretary of State's reasons for requiring modifications (unless it would be against the interests of national security, public safety or relations with the government of a country outside the UK i.e. foreign relations). Ofcom must comply with any direction and submit a revised Code as soon as reasonably practicable. When the Secretary of State is satisfied that no further modifications to the draft are required, the draft must be laid before Parliament.

A12.31 If a draft Code has been laid before Parliament following a direction and modifications under section 44(1), (2) or 3(b) of the Act then the affirmative procedure applies.[51] If a draft terrorism or CSEA Code has been the subject of a direction and modifications under section 44(3)(a), (4) or (5) then the negative procedure applies.[52]

[47] The Act, section 43.
[48] See sections 45(5) and (6) of the Act.
[49] See section 45(4) of the Act.
[50] The Act, section 44.
[51] Which is set out in section 45(4) of the Act.
[52] Which is set out in section 45(5) of the Act.

A12.32 Ofcom must publish each Code (or amendments to a Code) within three days of when it is issued.[53] Where we withdraw a Code of Practice, we must publish a notice to that effect.[54]

A12.33 We must keep each Code we publish under review.[55] The Secretary of State can require us to review a terrorism or CSEA Code for reasons of national security or public safety, and we must carry out such a review as soon as reasonably practicable. We must then make any necessary changes to the Code, or if we consider no changes are required, submit a statement to the Secretary of State explaining why.[56]

A12.34 Subject to the Secretary of State's approval, Ofcom may make minor amendments to a Code without consultation or laying the amendments before Parliament.[57]

A12.35 The safety duties apply to providers from the day on which the first relevant Code comes into force.[58]

Guidance

A12.36 The Act sets out various procedural requirements relating to the other forms of guidance that Ofcom is required to produce.

A12.37 In relation to the Illegal Content Judgements Guidance,[59] Ofcom is required to consult before producing the guidance (or revised or replacement guidance) and publish the guidance.

A12.38 Ofcom must also publish the Risk Profiles prepared under section 98 and from time to time review and revise the risk assessments and Risk Profiles so as to keep them up to date. Ofcom is further required to consult the Information Commissioner before producing our guidance (or revised or replacement guidance) about risk assessments under section 99. We must revise this guidance from time to time in response to further risk assessments under section 98 or to revisions of the Risk Profiles. Ofcom must publish this guidance.

A12.39 Under section 52, Ofcom must produce guidance for providers to assist them in complying with their duties set out in sections 23 or 34 regarding record-keeping and review and section 36 regarding children's access assessments. Ofcom must also produce guidance for Category 1 services relating to their duties set out in section 14 (assessments related to the adult user empowerment duty set out in section 15(2)) and section 18 (news publisher content). Ofcom must consult the Information Commissioner before producing this guidance (expect for the news publisher content guidance) and publish the guidance.

A12.40 In relation to Ofcom's guidance about enforcement action,[60] we must consult the Secretary of State, the Information Commissioner and such other persons we consider appropriate before producing the guidance (or revised or replacement guidance) and publish the guidance.

A12.41 Schedule 13 refers to Ofcom's penalty guidelines issued under s. 392 of the CA 2003 insofar as they are relevant to penalties imposed under the Act. Ofcom must consult on these

[53] The Act, section 46.
[54] The Act, section 46(3).
[55] The Act, section 47.
[56] The Act, section 47(3).
[57] The Act, section 48.
[58] The Act, section 51.
[59] The Act, section 193.
[60] The Act, section 151.

guidelines, in particular the Secretary of State, and publish the guidelines in a way we consider appropriate for bringing them to the attention of persons who are likely to be affected by them. The penalty guidelines may be included in the same document as the enforcement guidance.[61]

Impact assessments

A12.42 Impact assessments provide a valuable way of assessing the options for regulation and showing why the chosen option(s) was preferred. They form part of best practice policy making. This is reflected in section 7 of the CA 2003,[62] which means that Ofcom generally must carry out impact assessments in cases where it appears to us our proposals are important. Proposals that are important for the purposes of this section include preparing (or amending) a Code of Practice under section 41 of the Act; proposals which would be likely to have a significant effect on businesses or the general public; or where there is a major change in Ofcom's activities. As a matter of policy, Ofcom is committed to carrying out impact assessments in the great majority of our policy decisions. Our impact assessment guidance sets out our general approach to how we assess and present the impact of our proposed decisions.[63]

A12.43 As set out in section 7(5) of the CA 2003, Ofcom has discretion as to the substance and form of an impact assessment, and this will depend on the particular proposals being made. However, impact assessments which relate to proposals about Codes specifically or anything else for the purposes of the carrying out of Ofcom's online safety functions under the Act must include an assessment of the likely impact of implementing the proposal on small and micro businesses.[64]

Other relevant powers and duties

A12.44 There are several other duties in the Act which are relevant to the proposals covered in this consultation.

A12.45 As referred to above, section 66 contains a requirement to report CSEA content to the NCA. Specifically, this section requires:

a) a UK provider of user-to-user services to use systems and processes which secure (so far as possible) that the provider reports all detected and unreported CSEA content present on the service to the NCA (a non-UK provider of a user-to-user service must only do so in relation to UK-linked CSEA content); and
b) a UK provider of a search service to use systems and processes which secure (so far as possible) that the provider reports all detected and unreported CSEA content present on websites or databases capable of being searched by the search engine to the NCA (a non-UK provider of a search service must only do so in relation to UK-linked CSEA content).

A12.46 The duties on providers of search services apply to providers of combined services in relation to the search engine of the service. Providers' reports under this section must meet

[61] The Act, section 151(6).
[62] As amended by section 93 of the Act.
[63] Ofcom's Impact assessment guidance.
[64] The Act, section 93.

the requirements set out in regulations made by the Secretary of State under section 67 of the Act, including in relation to time frames.

A12.47 Section 64 of the Act contains requirements relating to user identity verification. Providers of Category 1 services must offer all adult users of the service in the UK the option to verify their identity. The provider must also include clear and accessible provisions in the terms of service explaining how the verification process works. Ofcom must issue guidance for providers of Category 1 services to assist them in complying with this duty.[65]

[65] The Act, section 65.

A12. Duties of Providers and Ofcom in relation to illegal content (Part B)

A12.1 This Annex sets out the duties relating to illegal harms, as they apply to providers of user-to-user services ('U2U services'); providers of search services; and to Ofcom, and which are relevant to this consultation. It should be read with reference to Annex 12.A, which sets out an overview of the Online Safety Act and Ofcom's powers in relation to it.

A12.2 This Annex does not cover other duties set out in the Act (except where relevant to illegal harms). Therefore, duties relating to, for example, the protection of children; user empowerment; the protection of content of democratic importance, news publisher content or journalistic content; fraudulent advertising; and other provisions of the Act are outside the scope of this consultation and will be addressed separately.[66]

Provider duties in relation to illegal content

A12.3 As summarised in in Annex 12.a, the Act imposes "duties of care" on providers of regulated user-to-user services ('U2U services'); and providers of regulated search services in relation to, among other things, "illegal content" (defined under section 59 of the Act. See also Chapter 2 of this consultation). Under the Act, "illegal content" is defined as "content that amounts to a relevant offence".[67] For U2U services, some of the duties apply in relation to the use of the service in question for the commission or facilitation of the defined priority offences identified in the Act (see paragraphs 2.24-28 of Chapter 2).

A12.4 The duties in relation to illegal content are set out in detail below.

Providers of U2U Services

A12.5 Providers of U2U services are given specific duties under the Act in relation to illegal content. These "Illegal content duties" include: "Illegal content risk assessment duties";[68] and "Safety duties about illegal content".[69]

A12.6 Providers of U2U services are also subject to "additional duties" which are relevant, among other things, to illegal content. These additional duties are as follows:

a) "Duties about content reporting and complaints procedures", which include:

i) "Duties about content reporting",[70] and
ii) "Duties about complaints procedures"[71] and

[66] For further information about consultations regarding the Online Safety Act, see Ofcom's approach to implementing the Online Safety Act.
[67] The Act, Section 59.
[68] The Act, section 9.
[69] The Act, section 10.
[70] The Act, section 20.
[71] The Act, section 21.

b) so-called "Cross-cutting duties", which include:

iii) "Duties about freedom of expression and privacy";[72] and
iv) "Record-keeping and review duties".[73]

A12.7 These are set out in more detail below. Section 7 of the Act states that all providers of regulated U2U services must comply with these duties (and the other duties set out under section 7(2)).

Connection with the United Kingdom

A12.8 These duties only apply to:

a) the design, operation and use of the service in the United Kingdom, and
b) in the case of a duty that is expressed to apply in relation to users of a service, the design, operation and use of the service as it affects United Kingdom users of the service.[74]

Combined Services

A12.9 Where the U2U service is a combined service (i.e. providing both a regulated U2U and regulated search service), these duties will not apply to:

a) the search content of the service,
b) any other content that, following a search request, may be encountered as a result of subsequent interactions with internet services, or
c) anything relating to the design, operation or use of the search engine.[75]

A12.10 However, the duties of care that apply to regulated search services in relation to illegal content (see paragraphs A12.38-57 below), will still apply.

Illegal Content Duties

Illegal content risk assessment duties

A12.11 Providers of regulated U2U services have a duty to carry out a suitable and sufficient illegal content risk assessment[76] at the specific times set out in Schedule 3 to the Act[77].

[72] The Act, section 22.
[73] The Act, section 23.
[74] The Act, section 8(3).
[75] The Act, section 8(2).
[76] Section 9(2).
[77] The deadline for completing the first risk assessment depends on the day on which a provider of U2U services starts its operations. In particular:

i. U2U services that are already in operation at the outset of this regime, must complete their first illegal content risk assessment within a period of three months from the day on which Ofcom's risk assessment guidance ('RAG') is published;
ii. new U2U services that start operations after the RAG is published must complete their first illegal content risk assessment within a period of three months from the day on which they begin their new services; and
iii. existing services that become U2U services (having previously provided a different type of service) after the RAG is published must complete their first illegal content risk assessment within a period of three months from the day on which their services become a U2U service. See Schedule 3 to the Act.

A12.12 An illegal content risk assessment means an assessment of the following matters, taking into account the risk profiles that relate to the services of that kind:[78]

a) user base;
b) the level of risk of individuals who are users of the service encountering, by means of the service, (i) each kind of priority illegal content (with each kind separately assessed) and (ii) other illegal content, taking into account (in particular) algorithms used by the service, and how easily, quickly and widely content may be disseminated by means of the service;
c) the level of risk of the service being used for the commission and/or facilitation of a priority offence;
d) the level of risk of harm to individuals presented by illegal content of different kinds or by the use of the service for the commission and/or facilitation of a priority offence;
e) the level of risk of functionalities of the service facilitating the presence or dissemination of illegal content or the use of the service for the commission or facilitation of a priority offence, identifying and assessing those functionalities that present higher levels of risk;
f) the different ways in which the service is used, and the impact of such use on the level of risk of harm that may be suffered by individuals;
g) the nature, and severity, of the harm that may be suffered by individuals from the matters identified in accordance with -paragraph (b) to (f) above; and
h) how the design and operation of the service (including the business model, governance, use of proactive technology, measures to promote users' media literacy and safe use of the service, and other systems and processes) may reduce or increase the risks identified.

A12.13 A Provider of a U2U service **must take appropriate steps to keep an illegal content risk assessment up to date**, including when OFCOM makes a significant change to a relevant risk profile. [79]

A12.14 A Provider of a U2U service is under an obligation **to carry out a further suitable and sufficient illegal content risk assessment, before making any significant changes** to any aspect of a service's design or operation - this further illegal content risk assessment must relate to the impact of that proposed change.[80]

Safety duties about illegal content

A12.15 Providers of regulated U2U services have specific safety duties in relation to illegal content as set out under Section 10 of the Act. These duties are as follows:

a) **A duty, in relation to a service, to take or use proportionate measures** relating to the **design or operation of the service** to—

i) prevent individuals from encountering priority illegal content by means of the service,
ii) effectively mitigate and manage the risk of the service being used for the commission or facilitation of a priority offence, as identified in the most recent illegal content risk assessment of the service, and

[78] Section 9(5).
[79] Section 9(3).
[80] Section 9(4).

iii) effectively **mitigate and manage the risks of harm to individuals**, as identified in the most recent illegal content risk assessment of the service (see paragraph A12.12(g)).[81]

b) **A duty to operate a service using proportionate systems and processes** designed to—

i) minimise the length of time for which any priority illegal content is present;
ii) where the provider is alerted by a person to the presence of any illegal content, or becomes aware of it in any other way, swiftly take down such content.[82]

c) **A duty to include** the following provisions in the **terms of service**:[83]

i) **Provisions specifying how individuals are to be protected from illegal content**. In particular, the terms of service must address how the provider intends to comply with the duty above at paragraph A12.15(b).[84];[85]
ii) **Provisions giving information about any proactive technology (see paragraphs A12.79-82 below) used by a service** for the purpose of compliance with the duties set out at paragraphs A12.15(a) or (b) above. This includes setting out the kind of technology that is being used, when it is used, and how it works.[86]
iii) Such provisions must be **clear and accessible**.[87]

d) A duty to **apply the provisions of the terms of service referred to above in paragraph A.12.15(c) consistently**.[88]

A12.16 A Provider of a category 1 service will also have **a duty to summarise in the terms of service the findings of the most recent illegal content risk assessment** of a service (including as to levels of risk and as to nature, and severity, of potential harm to individuals).[89]

A12.17 The duties set out at paragraphs A12.15(a) & (b)[90] apply across all areas of the provider's U2U service, including the way it is designed, operated and used as well as content present on the service. Among other things, these duties require the provider of a service, if it is proportionate to do so, to take or use measures in the following areas:

a) regulatory compliance and risk management arrangements,
b) design of functionalities, algorithms and other features,
c) policies on terms of use,
d) policies on user access to the service or to particular content present on the service, including blocking users from accessing the service or particular content,
e) content moderation, including taking down content,
f) functionalities allowing users to control the content they encounter,
g) user support measures, and

---

[81] The Act, ss 10(2)(a)-(c).
[82] The Act, ss 10(3)(a)-(b).
[83] Terms of service is defined under section 237 of the Act as: "in relation to a user-to-user service, means all documents (whatever they are called) comprising the contract for use of the service (or of part of it) by United Kingdom users".
[84] The Act, ss 10(5).
[85] In relation to paragraph A12.15(b)(i), the provider must specifically address terrorism content, CSEA content, and other priority illegal content.
[86] The Act, ss 10(7).
[87] The Act, ss 10(8).
[88] The Act, ss 10(6).
[89] The Act, ss10(9).
[90] I.e. The Act, ss 10(2)&(3).

h) staff policies and practices.[91]

A12.18 In determining what is **"proportionate"** for the purposes of the Safety Duties, the following factors, in particular, are relevant:

a) all the findings of the most recent illegal content risk assessment, including as to levels of risk and as to nature, and severity, of potential harm to individuals, and
b) the size and capacity of the provider of a service.[92]

Duties about content reporting and complaints procedures

A12.19 The Duties about content reporting and complaints procedures for providers of U2U services are contained in sections 20 and 21 of the Act.

Duties about content reporting

A12.20 All providers of regulated U2U services are required **to use systems and processes in the operation of their services which allow users and "affected persons" (see A12.22 below) to easily report certain types of content, depending on the kind of service**. For instance, such systems and processes must be put in place to enable users and affected persons to report "illegal content" on *all* U2U services.[93]

A12.21 For services that are likely to be accessed by children, the duty also applies in respect of content that is harmful to children.[94];[95]

A12.22 For the purposes of the duties about content reporting and complaints procedures (i.e. paragraphs A12.20-26), an **"affected person"** means a person, other than a user of the service in question, who is in the United Kingdom and who is: (a) the subject of the content, (b) a member of a class or group of people with a certain characteristic targeted by the content, (c) a parent of, or other adult with responsibility for, a child who is a user of the service or is the subject of the content, or (d) an adult providing assistance in using the service to another adult who requires such assistance, where that other adult is a user of the service or is the subject of the content.[96]

A12.23 In applying the content reporting duty, the cross-cutting duties will also be relevant.

Duties about complaints procedures

A12.24 There are two main duties in respect of complaints procedures which apply in relation to all regulated user-to-user services. These are:

a) **A duty to operate a complaints procedure**, in relation to a service, that:

i) allows for relevant kinds of complaint to be made (as set out below),
ii) provides for appropriate action to be taken by the provider of the service in response to complaints of a relevant kind, and

[91] The Act, ss 10(4).
[92] The Act, ss 10(10).
[93] The Act, ss 20(2)&(3).
[94] The Act, ss 20(2)&(4).
[95] Section 20(6) states that: "*a provider is only entitled to conclude that it is not possible for children to access a service, or a part of it, if age verification or age estimation is used on the service with the result that children are not normally able to access the service or that part of it.*"
[96] The Act, ss20(5).

iii) is easy to access, easy to use (including by children) and transparent.[97]

b) **A duty to include provisions in the terms of service which are easily accessible (including to children) specifying the policies and processes that govern the handling and resolution of complaints of a relevant kind**.[98]

A12.25 For all services, a relevant complaint will be:

a) complaints by users and affected persons about content present on a service which they consider to be illegal content;
b) complaints by users and affected persons (see definition at paragraph A12.22) if they consider that the provider is not complying with their: Illegal content duties (paragraph A12.11-18), the content reporting duty (paragraph A12.20), or either of the cross-cutting duties (paragraphs A12.27-33);
c) complaints by a user who has generated, uploaded or shared content on a service if that content is taken down on the basis that it is illegal content;
d) complaints by a user of a service if the provider has given a warning to the user, suspended or banned the user from using the service, or in any other way restricted the user's ability to use the service, as a result of content generated, uploaded or shared by the user which the provider considers to be illegal content;
e) complaints by a user who has generated, uploaded or shared content on a service if—

i) the use of proactive technology on the service results in that content being taken down or access to it being restricted, or given a lower priority or otherwise becoming less likely to be encountered by other users, and
ii) the user considers that the proactive technology has been used in a way not contemplated by, or in breach of, the terms of service (for example, by affecting content not of a kind specified in the terms of service as a kind of content in relation to which the technology would operate).[99]

A12.26 Services that are likely to be accessed by children and Category 1 services are required to provide for additional types of relevant complaint. For instance, if the service is likely to be accessed by children, then certain complaints regarding the provider's duties in relation to children's online safety will be relevant.[100] For Category 1 services, relevant complaints include complaints that they are not complying with their duties relating to user empowerment, content of democratic importance, news publisher content, journalistic content and freedom of expression and privacy[101]

Cross-cutting duties

A12.27 The Act also creates so-called "cross-cutting duties", which apply to regulated U2U services in relation to the performance of *other* duties under the Act. For instance, the freedom of expression and privacy duties are concerned with how "safety measures and policies" are introduced in relation to a regulated U2U service. These "safety measures and policies" refer to any measures or policies designed to secure compliance with the safety duties in respect of illegal content (section 10, paragraphs A12.15-18), and the duties about content reporting

[97] The Act, ss 21(2).
[98] The Act ss 21(3).
[99] The Act ss 21(4)(a)-(e).
[100] The Act, ss 21(5).
[101] The Act, ss 21(6).

(section 20, paragraphs A12.20-23) and complaints procedures (section 21, paragraphs A12.24-26), as well as other duties in relation to children's online safety (section 11), and user empowerment (section 15).

A12.28 In a similar vein, the record-keeping and review duties apply to the performance of the risk assessment duties under section 9 (and section 11); and other "relevant duties", including the illegal content duties (section 10), and content reporting (section 20) and complaints procedures (section 21).

A12.29 The cross-cutting duties for regulated U2U services are contained in sections 22 and 23 of the Act.

Duties about Freedom of Expression and Privacy

A12.30 All regulated U2U services will have the following duties when deciding on, and implementing, "safety measures and policies":

a) **a duty to have particular regard to the importance of protecting users' right to freedom of expression within the law**;[102] and
b) **a duty to have particular regard to the importance of protecting users from a breach of any statutory provision or rule of law concerning privacy that is relevant to the use or operation of a user-to-user service (including, but not limited to, any such provision or rule concerning the processing of personal data)**.[103]

A12.31 In addition, regulated U2U services which are also Category 1 services will have the following duties:

a) **A duty to carry out impact assessments:**

i) when deciding on safety measures and policies, to determine the impact that such measures or policies have on (i) users' right to freedom of expression within the law, and (ii) the privacy of users;[104] and
ii) to determine the impact that any adopted safety measures and policies have on (i) users' right to freedom of expression within the law, and (ii) the privacy of users.[105]

An impact assessment relating to a service must include a section which considers the impact of the safety measures and policies on the availability and treatment on the service of content which is news publisher content or journalistic content in relation to the service.

b) **A duty to keep an impact assessment up to date**, and **to publish impact assessments**.[106]
c) **A duty to specify in a publicly available statement the positive steps that the provider has taken in response to an impact assessment to**— (i) protect users' right to freedom of expression within the law, and (ii) protect the privacy of users.[107]

Record-keeping and review duties

A12.32 All regulated U2U services will have the following duties:

[102] The Act, ss 22(2).
[103] The Act, ss 22(3).
[104] The Act, ss 22(4)(a).
[105] The Act, ss 22(4)(b).
[106] The Act, s 22(6).
[107] The Act, s 22(7).

a) **A duty to make and keep a written record**, in an easily understandable form, **of every risk assessment** under section 9 (Illegal Content Risk assessment duties) or 11 (Children's Risk Assessments).[108]

b) **A duty to make and keep a written record of any measures taken or in use to comply with a relevant duty** which— (a) are described in a Code of Practice and recommended for the purpose of compliance with the duty in question, and (b) apply in relation to the provider and the service in question. Such measures are referred to as "applicable measures in a Code of Practice".[109]

c) If **alternative measures** (see paragraph A12.33 below) have been taken or are in use to comply with a relevant duty, **a duty to make and keep a written record containing the following information**—

i) the applicable measures in a Code of Practice that have not been taken or are not in use,

ii) the alternative measures that have been taken or are in use,

iii) how those alternative measures amount to compliance with the duty in question, and

iv) how the provider has had regard to the importance of protecting the right of users to freedom of expression within the law, and protecting the privacy of users in taking or using alternative measures.[110];[111]

d) **A duty to review compliance with the relevant duties in relation to a service**— (a)regularly, and (b) as soon as reasonably practicable after making any significant change to any aspect of the design or operation of the service.[112]

A12.33 'Alternative measures' means measures other than measures which are (in relation to the provider and the service in question) applicable measures in a Code of Practice. If alternative measures have been taken or are in use to comply with the safety duties about illegal content (as at paragraphs A12.15-18 above),[113] or a duty set out in section 11(2) or (3)) of the Act (safety duties protecting children), these records must also indicate whether such measures have been taken or are in use in every area listed at paragraphs A12.17(a)-(h) above or 11(5) (concerning safety duties protecting children), as the case may be, in relation to which there are applicable measures in a Code of Practice (see paragraphs A12.67-83 below).

Providers of Search Services

A12.34 Providers of regulated search services are also given specific duties under the Act in relation to illegal content. These "Illegal content duties for all search services" include: "Illegal content risk assessment duties";[114] and "Safety duties about illegal content".[115]

[108] The Act, ss 23(2).
[109] The Act, ss 23(3).
[110] The Act, ss 23(4).
[111] The Act, ss 49(5).
[112] The Act, ss 23(6).
[113] The Act, ss 10(2) or (3).
[114] The Act, ss 26.
[115] The Act, ss 27.

A12.35 Providers of regulated search services are also subject to additional duties which are relevant to illegal content, but also apply to other types of content and in respect of other regulatory requirements as set out under the Act. These are:

a) "Duties about content reporting and complaints procedures", which include:

i) The "Duty about content reporting", [116] and
ii) "Duties about complaints procedures";[117] and

b) the "Cross-cutting duties", which include:

iii) "Duties about freedom of expression and privacy";[118] and
iv) "Record-keeping and review duties".[119]

A12.36 The Illegal content duties for all search services; Duties about content reporting and complaints procedures; and the Cross-cutting duties that apply to providers of search services are set out in more detail below.

A12.37 These duties only apply to:

a) the search content of the service,
b) the design, operation and use of the search engine in the United Kingdom, and
c) in the case of a duty that is expressed to apply in relation to users of a service, the design, operation and use of the search engine as it affects United Kingdom users of the service.[120]

Illegal content duties for all search services

Illegal content risk assessment duties

A12.38 Providers of regulated search services have a duty to carry out a suitable and sufficient illegal content risk assessment[121] at the times set out in Schedule 3 to the Act.[122]

A12.39 An illegal content risk assessment of a service means an assessment of the following matters, taking into account the risk profile that relates to the service of that kind-

---

[116] The Act, s 31.
[117] The Act, s 32.
[118] The Act, s 33.
[119] The Act, s 34.
[120] The Act, section 25.
[121] section 26(2)
[122] The deadline for completing the first risk assessment depends on the day on which a search service's provider starts its operations. In particular:

i. search services that are already in operation at the outset of this regime must complete their first illegal content risk assessment within a period of three months from the day on which Ofcom's risk assessment guidance ('RAG') is published;
ii. new search services that start operations after the RAG is published must complete their first illegal content risk assessment within a period of three months from the day on which they begin their new services; and
iii. existing services that become search services (having previously provided a different type of service) after the RAG is published must complete their first illegal content risk assessment within a period of three months from the day on which their services become a U2U service.

See Schedule 3 to the Act.

a) the level of risk of individuals who are users of the service encountering search content of the following kind: (i) each kind of priority illegal content (with each kind separately assessed) and (ii) other illegal content, taking into account (in particular) risks presented by algorithms used by the service, and the way that the service indexes, organises, and presents search results;
b) the level of risk of functionalities of the service facilitating individuals encountering search content that is illegal content, identifying and assessing those functionalities that present higher levels of risks;
c) the nature, and severity, of the harm that might be suffered by individuals from the matters identified in accordance with paragraphs (a) and (b) above; and;
d) how the design and operation of the service (including the business model, governance, use of proactive technology, measures to promote users' media literacy and safe use of the service, and other systems and processes) may reduce or increase the risks identified.[123]

A12.40 After completing the first illegal content risk assessment, providers of regulated search services are under **a duty to take appropriate steps to keep an illegal content risk assessment up to date**, including when Ofcom make any significant change to a risk profile that relates to the services of the kind in question.[124]

A12.41 Before making any significant change to any aspect of a service's design or operation, providers of regulated search services are under a duty to carry out a further suitable and sufficient illegal content risk assessment relating to the impacts of the proposed change[125].

Safety duties about illegal content

A12.42 Providers of regulated search services have specific Safety duties in relation to illegal content as set out under section 27 of the Act. These duties are as follows:

a) **A duty**, in relation to a service, **to take or use proportionate measures relating to the design or operation of the service to effectively mitigate and manage the risks of harm to individuals, as identified in the most recent illegal content risk assessment of the service** (see paragraphs A12.38-41 above).[126]
b) **A duty to operate a service using proportionate systems and processes designed to minimise the risk of individuals encountering search content** of the following kinds—
   i) **priority illegal content**; and
   ii) **other illegal content** that the provider knows about (having been alerted to it by another person or become aware of it in any other way).[127]
c) **A duty to include provisions in a publicly available statement specifying how individuals are to be protected from search content that is illegal content**.[128]
d) **A duty to apply the provisions of the statement referred to at paragraph A12.42(c) above consistently**.[129]

[123] The Act, ss 26(5)(a)-(d).
[124] The Act, ss 26(3).
[125] The Act, ss 26(4).
[126] The Act, ss 27(2).
[127] The Act, ss 27(3).
[128] The Act, ss 27(5).
[129] The Act, ss 27(6).

e) **A duty to include provisions in a publicly available statement giving information about any proactive technology** (see paragraphs A12.79-82 below) used by a service for the purpose of compliance with a duty set out in sections 27(2) or (3) (paragraphs A12.42(a) or (b) above) (including the kind of technology, when it is used, and how it works).[130]

f) **A duty to ensure that the provisions of the publicly available statement** referred to in sections 27(5) and (7) (paragraphs A12.42(c)&(e) above) are **clear and accessible.**[131]

A12.43 The duties set out in paragraphs A12.42(a)-(b) above apply across all areas of a service, including the way the search engine is designed, operated and used as well as search content of the service. Among other things, these duties require the provider of a service to take or use measures in the following areas, if it is proportionate to do so:

a) regulatory compliance and risk management arrangements,
b) design of functionalities, algorithms and other features relating to the search engine,
c) functionalities allowing users to control the content they encounter in search results,
d) content prioritisation,
e) user support measures, and
f) staff policies and practices.[132]

A12.44 In determining what is '**proportionate**' for the purposes of the Safety Duties for Search Services, the following factors, in particular, are relevant:

a) all the findings of the most recent illegal content risk assessment (including as to levels of risk and as to nature, and severity, of potential harm to individuals), and
b) the size and capacity of the provider of a service.[133]

Duties about content reporting and complaints procedures

Duty about content reporting

A12.45 All providers of regulated search services are required **to operate a service using systems and processes that allow users and 'affected persons' to easily report certain types of search content, depending on the type of service**.[134] For instance, such systems and processes must be put in place to enable users and affected persons to report 'illegal content' on *all* of the search service.[135]

A12.46 For services that are likely to be accessed by children, the duty also applies in respect of content that is harmful to children.[136]

A12.47 For the purposes of the duties about content reporting and complaints procedures (i.e. paragraphs A12.45-51), an "affected person" has the same definition as for U2U services (see paragraphs A12.22 above).

Duties about complaints procedures

A12.48 There are two main duties in respect of complaints procedures which apply in relation to all regulated search services. These are as follows:

---

[130] The Act, ss 27(7).
[131] The Act, ss 27(8).
[132] The Act, ss 27(4).
[133] The Act, ss 27(10).
[134] The Act, ss 31(2).
[135] The Act, ss 31(3).
[136] The Act, ss 31(4).

a) A **duty to operate a complaints procedure** in relation to a service that—

i) allows for relevant kinds of complaint to be made (as set out below),
ii) provides for appropriate action to be taken by the provider of the service in response to complaints of a relevant kind, and
iii) is easy to access, easy to use (including by children) and transparent.[137]

b) **A duty to make the policies and processes that govern the handling and resolution of complaints of a relevant kind publicly available and easily accessible (including to children).**[138]

A12.49 Relevant complaints in relation to a regulated search service are:

a) complaints by users and affected persons (see paragraph A12.22 above) about search content which they consider to be illegal content;
b) complaints by users and affected persons if they consider that the provider is not complying with their illegal content duties (paragraph A12.38-44), content reporting duties (paragraphs A12.45-47), or freedom of expression and privacy (see paragraph A12.55);
c) complaints by an interested person if the provider of a search service takes or uses measures in order to comply with their safety duties (paragraphs A12.42-44) that result in content relating to that interested person no longer appearing in search results or being given a lower priority in search results;
d) complaints by an interested person if—

i) the use of proactive technology (see paragraphs A12.79-82 below) on a search service results in content relating to that interested person no longer appearing in search results or being given a lower priority in search results; and
ii) the interested person considers that the proactive technology has been used in a way not contemplated by, or in breach of, the provider's policies on its use (for example, by affecting content not of a kind specified in those policies as a kind of content in relation to which the technology would operate).[139]

A12.50 A complaint may also be a relevant complaint in the specific context of the service that is being provided. For instance, if the service is likely to be accessed by children, then certain complaints regarding the provider's duties in relation to children's online safety will be relevant.[140]

A12.51 For the purposes of the duties about complaints procedures for regulated search services, an 'interested person' means a person that is responsible for a website or database capable of being searched by the search engine, provided that—

a) in the case of an individual, the individual is in the United Kingdom;
b) in the case of an entity, the entity is incorporated or formed under the law of any part of the United Kingdom.[141]

[137] The Act, ss 32(2)(a)-(c).
[138] The Act, ss 32(3).
[139] The Act, ss 32(4)(a)-(d).
[140] See the Act, ss 32(5)(a)-(d).
[141] The Act, ss 32(6) & 228(7).

Cross-cutting duties

A12.52 The Act also creates 'cross-cutting' duties which apply to regulated search services in relation to the performance of other duties under the Act. For instance, the duties about freedom of expression and privacy are concerned with how "safety measures and policies" are introduced in relation to a regulated search service. These "safety measures and policies" refer to any measures or policies designed to secure compliance with the safety duties about illegal content (section 27, paragraphs A12.42-4 above), and the duty about content reporting (section 31, paragraphs A12.45-47 above), and duties about complaints procedures (section 32, paragraphs A12.48-51 above), as well as other duties in relation to children's online safety (section 29 – these duties are beyond the scope of this consultation).

A12.53 In a similar vein, the record-keeping and review duties apply to the performance of the risk assessment duties under section 26 (paragraph A12.38-41) and section 28 (Children's risk assessment duties); and other "relevant (duties)", including the safety duties in respect of illegal content (paragraphs A12.42-44 above), and content reporting and complaints procedures (see sections 31 and 32, paragraphs A12.45-51).

A12.54 The cross-cutting duties for regulated search services are set out in sections 33 and 34 of the Act.

Duties about freedom of expression and privacy

A12.55 All regulated search services will have the following duties when deciding on, and implementing, "safety measures and policies" (see above):

a) **a duty to have particular regard to the importance of protecting the rights of users and interested persons to freedom of expression within the law**;[142] and
b) **a duty to have particular regard to the importance of protecting users from a breach of any statutory provision or rule of law concerning privacy** that is relevant to the use or operation of a search service (including, but not limited to, any such provision or rule concerning the processing of personal data).[143]

Record-keeping and review duties

A12.56 All regulated search services will have the following duties:

a) **A duty to make and keep a written record, in an easily understandable form, of every risk assessment** made under section 26 (see paragraphs A12.38-41 above) or 28 (children's risk assessment duties).[144]
b) **A duty to make and keep a written record of any measures taken or in use** to comply with a relevant duty (see paragraph A12.56 above) which—

i) are described in a Code of Practice and recommended for the purpose of compliance with the duty in question, and
ii) apply in relation to the provider and the service in question. In this section such measures are referred to as "applicable measures in a code of practice".[145]

c) If alternative measures have been taken or are in use to comply with a relevant duty**, a duty to make and keep a written record containing the following information—**

[142] The Act, ss 33(2).
[143] The Act, ss 33(3).
[144] The Act, ss 34(2).
[145] The Act, ss 34(3).

iii) the applicable measures in a Code of Practice that have not been taken or are not in use,
iv) the alternative measures that have been taken or are in use,
v) how those alternative measures amount to compliance with the duty in question, and
vi) how the provider has had regard to the importance of protecting the right of users and interested persons to freedom of expression within the law, and protecting the privacy of users in taking or using alternative measures (i.e. under section 49(5)).[146]

If alternative measures have been taken or are in use to comply with the Safety duties about illegal content (specifically sections 27(2) or (3), as at paragraphs 42(a)-(b) above), or a duty set out in section 29(2) or (3) of the Act (Safety duties protecting children), this record must also indicate whether such measures have been taken or are in use in every area listed at paragraphs A12.43(a)-(f) above or section 29(4) (concerning Safety duties protecting children) (as the case may be) in relation to which there are applicable measures in a Code of Practice (see paragraphs A12.68-83).[147]

d) **A duty to review compliance with the relevant duties in relation to a service**— regularly, and as soon as reasonably practicable after making any significant change to any aspect of the design or operation of the service.[148]

A12.57 Ofcom may provide that particular descriptions of providers of search services are exempt from any or all of the record-keeping and review duties, and must publish details of any exemption.[149]

Ofcom's duties in relation to illegal content

A12.58 As set out in Chapter 2, Volume 1 of this consultation, the Act gives specific duties to Ofcom in relation to illegal content. These are set out below.

Ofcom sector risk assessment

A12.59 Ofcom is under a duty to carry out a risk assessment to identify and assess the risks of harm to individuals in the UK caused by:

a) illegal content on U2U services and by the use of U2U services for the commission and/or facilitation of priority offences (the ''illegality risks");
b) illegal content that appears to individuals in search results encountered on search services ("risk of harm from illegal content" and, together with the "illegality risks", the "risks of harm").[150]

A12.60 It has a discretion whether to combine these or consider them separately (also with the risk of harm to children under section 98(1)(c)).[151]

A12.61 Ofcom's risk assessment must, among other things, identify the characteristics of U2U and search services (which include functionalities, user base, business model and governance,

[146] The Act, ss 34(4)(a)-(d).
[147] The Act, ss 34(5).
[148] The Act, ss 34(6)(a)&(b).
[149] The Act, ss 34(7).
[150] The Act, ss 98(1)(a)&(b).
[151] The Act, ss 98(3).

and other systems and processes) that are relevant to the risks of harm and assess the impact of these characteristics on the risks of harm.[152] Please see Chapter 8, Volume 2, of this consultation.

Register of Risks and Risk Profiles

A12.62 Ofcom must prepare and publish a register of risks that reflects the findings of its risk assessments (the 'Register of Risks'). The Register of Risks must be prepared as soon as reasonably practicable after completion of the risk assessments.[153]

A12.63 Further to the Register of Risks, after completing its risk assessments, Ofcom must prepare and publish Risk Profiles for U2U services and search services that relate to each risk of harm, as applicable (the 'Risk Profiles'). In preparing the Risk Profiles, Ofcom can group U2U services and search services as appropriate and having regard to (i) the characteristics of the services and (ii) the risk levels and other matters identified in the risk assessment.[154]

A12.64 Ofcom must review and revise the risk assessments and the Risk Profiles from time to time to keep them up to date.[155] Please see Chapter 6, Volume 2 of this consultation.

Risk assessment guidance for services

A12.65 Ofcom must prepare and publish guidance to help U2U services and search services comply with their duties to prepare illegal content risk assessments under sections 9 and 26 respectively (the 'Risk Assessment Guidance' or 'RAG') (please refer to paragraphs A12.59-67).[156] Please refer to Annex 5.

A12.66 Ofcom must prepare the RAG as soon as reasonably practicable after having published the risk profiles relating to the risks of harm.

A12.67 Ofcom must revise and publish an updated RAG when it carries out a new risk assessment and/or revises the risk profiles.[157]

"Illegal" Codes for U2U and Search

Ofcom's duty to prepare and issue Codes of Practice in relation to illegal content

A12.68 Ofcom must prepare and issue Codes of Practice for providers of regulated U2U services and providers of regulated search services. The Codes of Practice must describe the measures Ofcom recommends these providers take for the purposes of complying with:

a) their respective safety duties in respect of illegal content (see sections 10 and 27, paragraphs A12.15-18 and A12.42-44), so far as they relate to:

i) Terrorism content or offences, as set out in Schedule 5 of the Act;[158] and
ii) CSEA content or offences, as set out in Schedule 6 of the Act;[159] and

b) the relevant duties (except to the extent they overlap with paragraphs A.12.68(a)(i) and (ii) above).[160] These include: the safety duties in respect of illegal content (see

[152] The Act, ss 98(2)&(11).
[153] The Act, s 98(4).
[154] The Act, ss 98(5)-(7).
[155] The Act, s 98(8).
[156] The Act, ss 99(1)&(2).
[157] The Act, s 99(5).
[158] The Act, s 41(1).
[159] The Act, s 41(2).
[160] The Act, s 41(3).

paragraphs A12.15-18 and A12.42-44); content reporting duties (see paragraphs A12.20-23 and A12.45-57); and complaints procedure duties (see paragraphs A12.27-33, and A12.48-51).[161]

A12.69 Schedule 4 of the Act sets out general principles and online safety objectives which the Codes must follow, as well as what content must be included. These are set out below.

General Principles

A12.70 In preparing a draft Code, Ofcom must consider the appropriateness of provisions of the Code to different kinds and sizes of U2U and search services, and to providers of differing sizes and capacities (paragraph 1 of Schedule 4). It must also have regard to the following principles:

a) providers of U2U and search services must be able to understand which provisions of the Code of Practice apply in relation to a particular service they provide;
b) the measures described in the Code of Practice must be sufficiently clear, and at a sufficiently detailed level, that providers understand what those measures entail in practice;
c) the measures described in the Code of Practice must be proportionate and technically feasible: measures that are proportionate or technically feasible for providers of a certain size or capacity, or for services of a certain kind or size, may not be proportionate or technically feasible for providers of a different size or capacity or for services of a different kind or size;
d) the measures described in the Code of Practice that apply in relation to U2U and search service providers of various kinds and sizes must be proportionate to Ofcom's assessment of the risk of harm presented by services of that kind or size (see paragraphs A12.59-61 above).[162]

Online Safety Objectives

A12.71 Ofcom must ensure that any measures described in the Codes are compatible with the pursuit of the online safety objectives.[163]

A12.72 For **U2U services**, these are:

a) That a service should be designed and operated in such a way that—
   i) the systems and processes for regulatory compliance and risk management are effective and proportionate to the kind and size of service,
   ii) the systems and processes are appropriate to deal with the number of users of the service and its user base,
   iii) UK users (including children) are made aware of, and can understand, the terms of service,
   iv) there are adequate systems and processes to support United Kingdom users, (v)(in the case of a Category 1 service) users are offered options to increase their control over the content they encounter and the users they interact with,

[161] The Act, s 41(10).
[162] The Act, Sch 4, subparas 2(a)-(d).
[163] The Act, Sch 4, para 3.

v) the service provides a higher standard of protection for children than for adults, (vii)the different needs of children at different ages are taken into account, (viii)there are adequate controls over access to the service by adults, and

vi) there are adequate controls over access to, and use of, the service by children, taking into account use of the service by, and impact on, children in different age groups; and

b) that a service should be designed and operated so as to protect individual UK users from harm, including with regard to—

vii) algorithms used by the service,

viii) functionalities of the service, and

ix) other features relating to the operation of the service.[164]

A12.73 For **search services**, these are:

a) That a service should be designed and operated in such a way that—

i) the systems and processes for regulatory compliance and risk management are effective and proportionate to the kind and size of service,

ii) the systems and processes are appropriate to deal with the number of users of the service and its user base,

iii) United Kingdom users (including children) are made aware of, and can understand, the publicly available statement referred to in relation to the Safety Duties (paragraph A12.42(c) above) and the safety duties protecting children (section 29)

iv) there are adequate systems and processes to support United Kingdom users,

v) the service provides a higher standard of protection for children than for adults, and

vi) the different needs of children at different ages are taken into account; and

b) that a service should be assessed to understand its use by, and impact on, children in different age groups; and

c) that a search engine should be designed and operated so as to protect individuals in the United Kingdom who are users of the service from harm, including with regard to—

vii) algorithms used by the search engine,

viii) functionalities relating to searches (such as a predictive search functionality), and

ix) the indexing, organisation and presentation of search results. [165]

A12.74 For **combined services**, these are:

a) That the online safety objectives that apply to U2U services (paragraphs A12.72(a)-(b) above) do not apply in relation to the search engine;

b) That the online safety objectives that apply to search services apply in relation to the search engine (and, accordingly, in this context, references to a search service include the search engine);

c) That the reference in a publicly available statement (as at paragraph A12.42(c) above) includes a reference to provisions of the terms of service which relate to the search engine. [166]

[164] The Act, Sch 4, subparas 4(a)-(b).

[165] The Act, Sch 4, subparas (5)(a)-(c).

[166] The Act, Sch 4, subparas 6(a)-(c).

A12.75 The Secretary of State may amend these objectives by way of regulations.[167]

Content of Codes of Practice

A12.76 The Act also sets out what type of measures must be included in the content of the Codes, and the principles in relation to which such measured should be designed. Such measures may only relate to the design or operation of the relevant service in the United Kingdom, or as it affects United Kingdom users of the service. In particular:

a) The Codes of Practice describing measures recommended for the purpose of compliance with the Safety Duties for providers of U2U services set out at paragraphs A12.15(a)&(b) above (i.e. in relation to taking proportionate measures relating to the design or operation of the service, or to operate a service using proportionate systems and processes), must include measures in each of the areas of a service listed at paragraphs A12.17(a)-(h).[168]
b) Codes of practice that describe measures recommended for the purpose of compliance with the Safety Duties about illegal content for providers of search services set out at paragraphs A12.42(a)&(b) (i.e. in relation to taking proportionate measures relating to the design or operation of the service, or to operate a service using proportionate systems and processes) must include measures in each of the areas of a service listed at paragraphs A12.43(a)-(f) above.[169] [170]

A12.77 Any measures described in a Code of Practice which are recommended for the purpose of compliance with any of the relevant duties must be designed in the light of the following principles:

a) the importance of protecting the right of users and (in the case of search services or combined services) interested persons to freedom of expression within the law, and
b) the importance of protecting the privacy of users.[171]

A12.78 Where appropriate, such measures must also incorporate safeguards for the protection of the matters mentioned in those principles.

Proactive technology

A12.79 If Ofcom considers it appropriate to do so, and in accordance with the general principles set out at paragraphs 1 and 2 of Schedule 4 (see paragraphs A12.70-74) and the principles set out at paragraph 10(2) of Schedule 4 (see paragraph A12.77), it may include in a Code of Practice a measure describing the use of a kind of technology. However, there are constraints on Ofcom's power to include a measure describing the use of "proactive technology" (a "proactive technology measure"). Section 231 defines "proactive technology" as consisting of three types of technology: content identification technology, user profiling technology, and behaviour identification technology (subject to certain exceptions). These are explained in greater detail below.

---

[167] The Act, Sch 4, para 7.
[168] The Act, Sch 4, subpara 9(1).
[169] i.e. the measures set out in section 27(4).
[170] The Act, Sch 4, subpara 9(3).
[171] This refers to protecting the privacy of users from a breach of any statutory provision or rule of law concerning privacy that is relevant to the use or operation of a U2U or search service (including any provisions concerning the processing of personal data), paragraph 10(4), Schedule 4.

A12.80 **Content identification technology** refers to technology, such as algorithms, keyword matching, image matching or image classification, which analyses content to assess whether it is content of a particular kind (for example, illegal content). Content identification technology is not regarded as proactive technology if it is used in response to a report from a user or other person about particular content.

A12.81 **User profiling technology** means technology which analyses (any or all of) relevant content (as defined in section 231(8)), user data, or metadata relating to relevant content or user data, for the purposes of building a profile of a user to assess characteristics such as age. However, technology which analyses data specifically provided by a user for the purposes of the provider verifying or estimating the user's age in order to decide whether to allow the user to access a service (or part of a service) or particular content, but which does not analyse any other data or content, is not regarded as user profiling technology.

A12.82 **Behaviour identification technology** means technology which analyses (any or all of) relevant content (as defined in section 231(8)), user data, or metadata relating to relevant content or user data, to assess a user's online behaviour or patterns of online behaviour (for example, to assess whether a user may be involved in, or be the victim of, illegal activity). But behaviour identification technology is not regarded as proactive technology if it is used in response to concerns identified by another person or an automated tool about a particular user.

A12.83 Ofcom has power to include a proactive technology measure in a Code of Practice for the purpose of compliance with the safety duties in relation to illegal content set out in sections 10(2) or (3) (for U2U services), or in sections 27(2) or (3) (for search services).[172] However, that power is subject to the following constraints:

a) A proactive technology measure may not recommend the use of technology which operates (or may operate) by analysing user-generated content communicated privately, or metadata relating to such content.[173]
b) A proactive technology measure may be included in a Code of Practice in relation to services of a particular kind or size only if Ofcom is satisfied that the use of the technology by such services would be proportionate to the risk of harm that the measure is designed to safeguard against (taking into account, in particular, Ofcom's risk profile relating to such services published under section 98, see paragraphs A12.63&64).[174]
c) In deciding whether to include a proactive technology measure in a Code of Practice, Ofcom must have regard to the degree of accuracy, effectiveness and lack of bias achieved by the technology in question. Ofcom may also refer in the Code of Practice to existing industry or technical standards for the technology (where they exist), or set out

[172] Paragraph 13(3) of Schedule 4 sets out that a proactive technology measure may also be recommended for the purpose of compliance with the children's online safety duties set out in section 12(2) or (3) (in relation to U2U services) or section 29(2) or (3) (in relation to search services), or for the purpose of compliance with the fraudulent advertising duties set out in section 38(1) or 39(1).
[173] See paragraph 13(4) of Schedule 4. For factors which Ofcom must particularly consider when deciding whether content is communicated "publicly" or "privately" by means of a user-to-user service for these purposes, see section 232.
[174] See paragraph 13(5) of Schedule 4.

principles in the Code of Practice designed to ensure that the technology or its use is (so far as possible) accurate, effective and free of bias.[175]

Relationship between provider duties and Ofcom's Codes of Practice

A12.84 Providers of a regulated U2U or search service who take or use the measures described in a Code of Practice which are recommended for the purpose of complying with a relevant duty will be treated as having complied with that relevant duty.[176] Further, providers who take or use the relevant recommended measures that incorporate safeguards to protect users' rights to freedom of expression within the law, and to protect the privacy of users, respectively, will be treated as having complied with the freedom of expression and privacy duties set out in sections 22(2)-(3), for U2U services, and sections 33(2)-(3), for search services, respectively.[177]

A12.85 Where a provider adopts an alternative measure to those described in a Code of Practice in order to comply with a relevant duty, it must have particular regard to the importance of: protecting the right of users and (in the case of search services) interested persons to freedom of expression within the law, and protecting the privacy of users.[178]

A12.86 When it is assessing whether a provider of a service is compliant with a relevant duty where that provider has adopted an alternative measure, Ofcom must consider the extent to which an alternative measure taken or in use by the provider extends across the relevant duties (i.e. under sections 10(4), or 29(4)), and, where appropriate, that it incorporates safeguards for the protection of the right of users and (in the case of search services) interested persons to freedom of expression within the law, and protection of the privacy of users.[179]

Effect of the Codes of Practice

A12.87 Failure to comply with a provision of a Code of Practice does not in itself make the provider liable to legal proceedings in a court or tribunal,[180] although the Code will be admissible in evidence in legal proceedings,[181] and any such court or tribunal must take a provision of the Code into account when determining a question which is relevant to that provision, as long as the question relates to a time when the provision was in force.[182] Similarly, Ofcom must take into account a provision of a Code of Practice when determining a question which is relevant to that provision, as long as the question relates to a time when the provision was in force.[183]

Illegal Content Judgements Guidance

A12.88 Providers of regulated U2U or search services complying with their duties as set out above will need to make judgments about whether content is content of a particular kind, on the basis of all relevant information reasonably available to them.[184] This includes decisions in

[175] See paragraph 13(6) of Schedule 4. This requirement does not apply to proactive technology which is a kind of age verification or age estimation technology: see paragraph 13(7) of Schedule 4.
[176] The Act, s 49(1).
[177] The Act, s 49(2)-(3).
[178] The Act, s 49(5).
[179] The Act, s 49(6).
[180] The Act, s 50(1).
[181] The Act. s 50(2).
[182] The Act, s 50(3).
[183] The Act, s 50(4).
[184] The Act, s 192.

relation to whether a provider has reasonable grounds to infer that content is content is illegal content, or illegal content of a particular kind.[185] Please see Chapter 2, Volume 1, for a more detailed definition of "illegal content".

A12.89 In order to make a judgement that content is illegal content, providers will need reasonable grounds to infer that all of the elements necessary for the commission of the offence, including the mental elements, are present or satisfied,[186] and that no defence to the offence may be successfully relied upon.[187]

A12.90 To assist providers in making these judgments in relation to illegal content, Ofcom must produce and publish Illegal Content Judgments Guidance ('ICJG').[188]

Enforcement guidance

A12.91 Ofcom must produce guidance for providers of regulated services about how it proposes to exercise its functions in relation to enforcement (these functions are set out at sections 130-150).[189] This guidance must give information about the factors Ofcom would consider it appropriate to take into account when taking, or considering taking, enforcement action relating to a provider's failure to comply with the different "enforceable requirements" set out in section 131 of the Act.[190] These include all of the duties set out above at paragraphs A12.5 – 57 above. The Guidance must also include provision explaining how Ofcom will take into account the impact or possible impact of such a failure on children when considering a failure to comply with the illegal content duties (sections 10 or 27), or any of the duties relating to children's online safety (sections 12 or 29) or children's access to provider pornographic content (section 81(2)).[191] Ofcom must have regard to this guidance when exercising their functions in relation to enforcement, or deciding whether to exercise them.[192]

Record keeping guidance

A12.92 Ofcom must produce guidance for providers of regulated U2U and search services to assist them in complying with their record-keeping and review duties (sections 23 (U2U) and 34 (search)) – paragraphs A12.32&33, and A12.56&57 above.[193]

Penalty guidelines

A12.93 Ofcom must prepare and publish a statement containing the guidelines they propose to follow in determining the amount of penalties imposed by them.[194]

---

[185] The Act, s 192(4)&(5).
[186] The Act, s 192(6)(a).
[187] The Act, s 192(6)(b).
[188] The Act, s 193.
[189] The Act, s 151(1).
[190] The Act, s 151(2).
[191] The Act, s 151(3).
[192] The Act, s 151(7).
[193] The Act, s 52(3).
[194] The Communications Act 2004, s 392(1). See also The Act, Sch 13, sub-para 2(5).

A13. Impact assessments

A13.1 This annex outlines our Equality Impact Assessment and Welsh language assessment.

Equality Impact Assessment

A13.2 We have given careful consideration to whether the proposals in this document will have a particular impact on persons sharing protected characteristics (including race, age, disability, sex, sexual orientation, gender reassignment, pregnancy and maternity, marriage and civil partnership and religion or belief in the UK and also dependents and political opinion in Northern Ireland), and in particular whether they may discriminate against such persons or impact on equality of opportunity or good relations. This assessment helps us comply with our duties under the Equality Act 2010 and the Northern Ireland Act 1998.

A13.3 We consider that some of our proposals would have a positive impact on certain groups. We consider that most of these impacts are likely to come from our Codes of Practice proposals. Specifically:

- **Terms of Service**: Our proposals that relate to comprehensibility of language may benefit those with protected characteristics which could affect their level of literacy. Benefits could accrue to younger users, those who may not have English as a first language (which can be associated with race) or those with relevant disabilities (e.g. Learning disability, visual impairment and motor impairment). We have also made specific proposals for the benefit of users of assistive technologies including: keyboard navigation; and screen reading technology.
- **User Complaints**: Our proposals to make complaints systems findable and accessible should help vulnerable users find and use them more easily. It should allow users with certain disabilities (e.g. learning disability, vision impairment and motor impairment) and some older users to easily access complaints processes. We have also made specific proposals for the benefit of users of assistive technologies including: keyboard navigation; and screen reading technology. Having an accessible and findable complaints process in turn should enable complaints about harm online occurring because of protected characteristics such as sex, sexual orientation, race or disability to be addressed appropriately.
- **Content and Search Moderation:** Our proposals for larger and riskier services to have content policies and appropriate training should in turn improve awareness of issues affecting groups with protected characteristics and the consistency of decision making in relation to them. We are proposing that services should have regard to the languages in which UK users encounter content when they resource their content moderation functions, which is likely to benefit speakers of languages other than English. This in turn may have positive impacts for those of nationalities other than English, who may be of many different races. The implementation of an effective content moderation function should improve outcomes for any group disproportionately subject to threats, abuse and harassment, which is likely to include most groups with protected characteristics.
- **CSAM Hash Matching, CSAM URL detection and Grooming mitigations**: Our proposals here will help to provide additional protections for children. As CSAM and child abuse more generally disproportionately affect women, it will also provide additional protections for them.

- **Service design user blocking and turning settings to private:** Our proposals to allow users to block specific pieces of content retroactively and turn their settings to private for people they do not know, should result in positive impacts for women, who tend to disproportionately face issues online such as harassment and stalking. It should also have positive impacts on people from different races, religions, sexual orientation, and those who have undergone gender reassignment, as they too tend to disproportionately experience certain types of abuse including hate speech (where relevant) and harassment.
- **Search**: Our proposals in relation to predictive search results are likely in our view to reduce the likelihood of users being prompted to run searches for hate speech, which in turn should benefit those groups with protected characteristics who tend to be targeted by such speech. Our proposals that search services should provide support information in response to searches for suicide should benefit all those at risk, which is likely to disproportionately include groups with protected characteristics.

A13.4 More generally, we consider that the work we are doing under the banner of "Promoting Compliance", will help increase the accessibility of our documentation, which should yield benefits to a range of equality groups (e.g. age, disability etc).

A13.5 At this stage, we do not envisage that our proposals would have a detrimental impact on any particular group of people.

A13.6 We recognise that it may be possible in some areas to do more to advance equality of opportunity and foster good relations between persons who share protected characteristics and persons who do not. We expect our evidence base and understanding to improve over time, and to be able to iterate our Codes of Practice. At this stage, in our view, our understanding of the costs and possible unintended consequences of seeking to do more is not sufficient.

Welsh Language Assessment

A13.7 The Welsh language has official status in Wales. To give effect to this, certain public bodies, including Ofcom, are required to comply with Welsh language standards.[195] Accordingly, we have considered:

- the potential impact of our policy proposals on opportunities for persons to use the Welsh language;
- the potential impact of our policy proposals on treating the Welsh language no less favourably than the English language; and
- how our proposals could be formulated so as to have, or increase, a positive impact; or not to have adverse effects or to decrease any adverse effects.

A13.8 Ofcom's powers and duties in relation to online safety regulation are set out in the Online Safety Act and must be exercised in accordance with our general duties under section 3 of the Communications Act 2003. In formulating our proposals in this Consultation, where relevant and to the extent we have discretion to do so in the exercise of our functions, we

[195] The Welsh language standards with which Ofcom is required to comply are available on our website.

have considered the potential impacts on opportunities to use Welsh and treating Welsh no less favourably than English: see in particular our proposals in relation to record-keeping.[196] More generally, we are proposing that services should have regard to the needs of their user base in considering what languages are needed for their content moderation, complaints handling, terms of service and publicly available statements. To this extent, we consider our proposals are likely to have positive effects or increased positive effects on opportunities to use Welsh and treating Welsh no less favourably than English.

What input do we want from stakeholders?

- Do you agree that our proposals as set out in Chapter 16 (reporting and complaints), and Chapter 10 and Annex 6 (record keeping) are likely to have positive, or more positive impacts on opportunities to use Welsh and treating Welsh no less favourably than English?
- If you disagree, please explain why, including how you consider these proposals could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.

[196] See Volume 3, Chapter 10, para 11.11(b) and Annex 7, paragraph A7.12.

A14. Further analysis on costs and benefits

A14.1 This annex provides further analysis which has been used to support our provisional conclusions for some of the measures we propose to include in our Illegal Content Codes of Practice ("Codes"). We outline:

a) Assumptions we have used to develop quantified cost estimates across a number of the measures;
b) In relation to our proposed measure for hash matching for child sexual abuse material (CSAM) (discussed in Chapter 14 on automated content moderation for U2U services), the additional analysis undertaken to support our provisional conclusions; and
c) In relation to our proposed measures for default settings for child users (discussed in Chapter 18 on default settings and user support for U2U services), the analysis underpinning our choice of option for targeting the proposed measures.

Assumptions on costs

A14.2 This annex describes some of the assumptions we have made on costs where these assumptions apply to many of the proposed measures. These assumptions are usually combined with other assumptions that are specific to each measure for determining the costs of measure in the chapters in the main body of the report. Any additional assumptions that are used in the cost analysis are described in the costs section of the relevant chapters.

Price Level

A14.3 All quantified estimates of costs or benefits are provided in 2022 prices, unless otherwise stated. We have used 2022 prices, as that is the year of the most recent Annual Survey of Hours and Earnings ('ASHE'), which we use to develop estimates for the labour cost required to implement some code measures.[197]

A14.4 When source data is not directly available in 2022 prices, then we inflate using the UK GDP deflator.[198]

Labour Costs

A14.5 To develop estimates for labour costs, we have estimated a salary range for three types of professions who are likely to develop and/or manage the systems and processes that in-scope services will need to have to comply with the regime. For the lower end of the range, we have used the ASHE 2022 gross median full-time earnings for the relevant occupation, which includes both base and incentive pay.[199]

---

[197] Office for National Statistics ("ONS"), 2022. *Annual Survey of Hours and Earnings ('ASHE'),* Table 14, 2022 provisional estimates.

[198] Using the GDP deflator is the approach recommended in HM Treasury's Green Book. The GDP deflator data from the ONS is available here.

[199] ASHE documentation does not explicitly state that gross salaries include bonuses, but our understanding is that the gross pay includes bonuses, tips and other payments.

A14.6 The three professions we have determined to be most relevant for developing our proposed measures, and their relevant Standard Occupational Classification ("SOC") 2020 references are:

a) We use the Programme and software development professionals salary (2134) from ASHE to estimate the cost of 'software engineer' time used when developing our cost estimates.
b) We use the Database administrators and web content technicians (3133)[200] salary from ASHE to estimate the cost of 'content moderator' time used when developing our cost estimates.
c) We use the Professional Occupations (2) estimate from ASHE to cover a range of professions that are employed at various online services and might be required to implement code measures. This could be legal employees, operations, product managers and so forth.

A14.7 We recognise that for some services, median UK wage rates may be too low. This may be especially the case for larger services based in the US. We also appreciate that the costs of hiring some types of staff, such as software engineers, may be considerably higher. To take account of this, we also include a higher estimate, which we have assumed is double the value of our lower estimate.

A14.8 Conversely, we are aware that some services may outsource some relevant work to locations where average pay is lower than the UK, which may reduce these costs. To the extent this is the case, our salary range may tend to overstate costs, making our analysis in support of our provisional conclusions on when measures are proportionate conservative.

A14.9 Table A14.1 shows the resulting low and high estimates we use for the three occupations.

Table A14.1: Gross Annual Wages Estimates:

Occupation	Gross Annual Wage Estimates (ASHE 2022)	
	Low	High
Software Engineer	£45,508	£91,016
Content Moderator	£30,461	£60,922
Professional Occupations	£41,604	£83,208

A14.10 We also assume **a 22% uplift** to the gross wage costs to account for non-wage labour costs, such as employers' National Insurance contributions.[201]

A14.11 When producing cost estimates for our measures, we have used resourcing estimates based on different time periods (e.g. days/months), depending what is appropriate for the

[200] This four-digit SOC 2020 code (unit group code 3133) includes occupations such as content, chat, web, and website moderators as well as other occupations such as database administrators and web content technicians. ONS, SOC 2020 Volume 2: the coding index and coding rules and conventions [accessed 29 September 2023]. The associated ONS spreadsheet can be found here: SOC 2020 Volume 2: the coding index.
[201] This is the non-wage uplift recommended by the Regulatory Policy Committee ("RPC"). Source: RPC, 2019. *RPC guidance note on 'implementation costs'*. It is also the uplift used by DSIT in its Impact Assessment for the Online Safety Bill.

particular measure. To help understand more clearly the unit labour costs used in each of these situations, Table A14.2 provides the daily labour costs[202] and Table A14.3 provides the monthly labour costs[203] that we have used. These figures include the 22% uplift mentioned above.

Table A14.2: Estimated daily labour cost

Occupation	Estimated Daily Labour Cost	
	Low	High
Software Engineer	£244	£488
Content Moderator	£163	£326
Professional Occupations	£223	£446

Table A14.3: Estimated monthly labour cost

Occupation	Estimated Monthly Labour Cost	
	Low	High
Software Engineer	£4,627	£9,253
Content Moderator	£3,097	£6,194
Professional Occupations	£4,230	£8,459

Maintenance Costs for System Changes

A14.12 Where system or other software changes associated with a proposed measure involve an initial cost, we assume there is also an ongoing annual maintenance cost of 25% of the initial cost. These ongoing costs reflect likely work required to ensure the system continues to operate as intended. We apply this assumption unless we have more specific information about the ongoing maintenance costs.

Further analysis on CSAM hash matching measure

A14.13 To support our provisional conclusion for our measure on CSAM hash matching contained within Chapter 14 on automated content moderation for U2U services, we have undertaken additional analysis to quantify the costs and benefits associated with this measure. The costs and benefits of the measure will vary by service due to many factors. Based on our current understanding, for the purpose of our analysis we assume that costs and benefits vary based

[202] The daily labour cost is estimated by increasing the annual salary by 22% and dividing by the number of working days in a year. We assume on average there are 228 working days in a year. This assumes people work 5 days a week and that there are 8 bank holidays and on average people take 25 days leave a year.
[203] The monthly labour cost is estimated by increasing the annual salary by 22% and dividing by the number of months in a year (12).

on two factors: a service's number of users and a service's risk of image-based CSAM. To align with our provisional conclusion in Chapter 14, we model three kinds of service based on their expected level of risk: medium-risk services (as would be determined by their risk assessment); high-risk services (also determined by their risk assessment), and file-storage and file-sharing services (for reasons outlined in paragraph 14.112 of Chapter 14). In this section we present the estimated costs and benefits for the smallest service of each kind that we are provisionally recommending the measure to.

A14.14 We have drawn on market prices, industry experts, and our own expertise to inform our analysis of costs. Our analysis of benefits was limited by a lack of evidence on the monetary value associated with removing CSAM from internet services. Monetised estimates are only available for the social cost of contact CSA, so we only estimate the benefit of removing CSAM from a service inasmuch as this results in a reduction in contact CSA. Due to data limitations, we do not quantify any other benefits associated with removing CSAM from a service, such as reduced re-victimisation and fewer people inadvertently viewing CSAM. Even when limiting our quantitative analysis to focus on the reduction in contact CSA, we find that this sole benefit could outweigh the direct costs of the measure for many services.

A14.15 This analysis supports our provisional conclusion by demonstrating that, even when only this single benefit is considered, the reduction of harm could be so significant that the direct costs of the measure are proportionate, including for the smallest services that we are provisionally recommending the measure for.

A14.16 However, we have limited evidence about the capacity of the hash-matching ecosystem to absorb a very significant increase in demand. We have therefore balanced the pressure that the proposed option would put on the relevant organisations when assessing which services it would be proportionate to recommend this measure to apply to, as discussed further in Chapter 14.

Estimating Costs

A14.17 We have quantified the following costs incurred by a service provider that implements the measure:

a) the one-off cost of building a hash-matching system ('build cost');
b) the on-going cost of maintaining a hash-matching system ('maintenance cost');
c) the on-going cost of software, hardware, and data ('tech cost'); and
d) the on-going cost of reviewing matches, moderating content, and reporting CSAM ('moderation cost').

A14.18 We understand that these costs are the most material costs associated with a service implementing the measure, but we recognise that there may be other costs to the service provider as well as to other organisations and individuals.

A14.19 Costs are calculated in terms of a range of plausible estimates to reflect uncertain and idiosyncratic factors affecting a service's cost profile. We refer to a service that could implement the measure at a cost equal to the lower bound of the range as having a 'low-cost profile'. Conversely, a service that with implementation costs at the top of the range is referred to as having a 'high-cost profile'. We would expect the costs for most services to lie within the range, but we recognise that there will be exceptions to this.

A14.20 Below, we explain how we have estimated the magnitude of these four costs:

a) **The one-off build cost** is estimated to range from £16,000 to £319,000. This assumes that the system takes between 2 to 18 months full-time work for a software engineer to build, matched by input from other professionals such as lawyers and product managers. [204] Services with a low-cost profile are assumed to pay median wages (see Table A14.1) and to take the minimum time required to build the system, consistent with the use of a third-party API-based solution.[205] Services with a high-cost profile are assumed to pay double the median wage and to take the equivalent of 18 months for engineers and other professionals to build the system.

b) **The annual maintenance cost** is estimated to range from £4,000 to £80,000 per year. Maintaining the system involves activities such as applying updates, adjusting parameters, and ingesting new hash lists. As explained in paragraph A14.12, we assume that annual maintenance costs are 25% of the initial cost required to build the system, which could include input from engineers and other professionals. As the build costs are calculated for services with a low-cost and high-cost profile, so too are the maintenance costs.

c) **The annual tech cost** is assumed to vary by the size of service. Our estimate draws on our own expertise, engagement with industry experts, and publicly available price points for safety technology solutions, including all-in-one software solutions.[206] Based on these sources we have assumed a minimum annual cost of £25,000 and have calibrated our model such that a service which reaches 25% of the UK's population will pay £275,000.[207] In general, we expect that larger services will pay more for software, hardware, and data because third-party organisations may base their price on the service's capacity to pay for the product and because larger services may opt to build in-house hash-matching systems that ingest multiple hash lists and involve bespoke software solutions.[208]

d) **The annual moderation cost** depends on the amount of positive matches that a service would discover on their service if they introduced hash matching. We estimate this based on the historical number of reports to NCMEC by services that already do hash matching. To allow us to generalise this number to services of any size, we assume that there is a relationship between a service's number of reports and number of users. Taking as our example a potentially high-risk service that already does hash matching, we calculate that this service made 0.0005 reports per user in 2022.[209] This ratio allows us to estimate the expected number of reports for similarly risky services of different sizes. For example, a service that has a reach of 700,000 UK users and makes 0.0005 reports per user would be expected to make 362 UK-based reports per year.[210] We use the example of another internet service, [**CONFIDENTIAL ✂**], to benchmark the cost of dealing with this much

---

[204] The time range is based on our own expertise and engagement with industry experts. This time range is conservative relative to the assumption made by the EU, that it takes 120 hours per year to build a system to detect known CSAM (see p218 of the EU's impact assessment). Source: European Commission, 2022. Proposal for a regulation laying down the rules to prevent and combat child sexual abuse: Impact assessment. [accessed 05 June 2023].

[205] Safety technology organisations offer API-based solutions which automatically check whether material on a service matches with any material on a CSAM hash list.

[206] IWF publish their membership fees and AWS publish price points for Thorn's Safer.

[207] We have assumed that the cost scales linearly with a service's user base.

[208] For example, some NGOs such as IWF and Thorn charge members depending on capacity to pay and other considerations.

[209] This is based on the number of reports by the service with CSAM pertaining to UK offenders or victims, relative to the number of UK users. The National Crime Agency (NCA) provided us with data on reports passed to them from NCMEC. CSAM data sourced from: [**CONFIDENTIAL ✂**]. Adult user numbers data sourced from: Ipsos, *Ipsos iris online audience measurement service*, January 2023, age: 18+, UK.

[210] We note that a report may include multiple images or videos.

CSAM, including with false positives and content that does not relate to the UK. The service told us that they employ [**CONFIDENTIAL ✂**] FTE content moderators just for CSAM. In the same year, the service made [**CONFIDENTIAL ✂**] reports of CSAM that pertained to UK offenders or victims, implying that they employed 0.002 moderators per such report. Applying this ratio, we would expect a service that makes, for example, 362 UK-based reports, to employ the full-time equivalent of 0.7 moderators.[211] Taken together, this implies that a service that a high-risk service that reaches 700,000 UK users would be expected to spend £21,000 to £61,000 on labour to review, moderate, and report CSAM. This calculation can be generalised to estimate the moderation costs for services with any number of users and different levels of risk.[212]

A14.21 Using the above assumptions, we can calculate the one-off and on-going costs for services with different numbers of users and levels of risk. To align with our provisional conclusion, we estimate the costs for three hypothetical services consistent with the user and risk thresholds set out in Chapter 14: a medium-risk service that reaches 7 million UK users; a high-risk service that reaches 700,000 UK users; and a file-storage and file-sharing service that reaches 70,000 UK users. We have applied the low-cost and high-cost profiles to each hypothetical service to form a range. However, we consider it very unlikely that a service with a small user base would have a high-cost profile or that a service with a large user base would have a low-cost profile. For example, services with a large user base generally have more elaborate organisational structures and complex product portfolios that result in higher build costs. Nevertheless, we have modelled costs for services with a low-cost and large user base, as well as for services with a high-cost profile and small user base, even though we consider both scenarios unlikely.

Table A14.4. Illustrative costs of three hypothetical services implementing the measure

		File-storage and file-sharing service that reaches 70,000 UK users	High risk service that reaches 700,000 UK users	Medium-risk service that reaches 7,000,000 UK users
Build costs (one-off)	Low-cost profile	£16,000	£16,000	£16,000
	High-cost profile	£319,000	£319,000	£319,000
On-going costs (annual)	Low-cost profile	£31,000	£59,000	£145,000
	High-cost profile	£110,000	£176,000	£254,000

[211] We add or subtract 20% from the central FTE estimate in order to estimate FTE under the high and low-cost profile, respectively.
[212] We have modelled services with different risk levels by using the ratio of reports to users from three different internet services that could be considered reflective of high-risk, medium-risk, and file-storage and file-sharing services.

Source: Ofcom analysis, various sources

Estimating Benefits

A14.22 The direct impact of hash matching is to identify known CSAM on a service which can then be removed. Removing CSAM has many benefits, such as reduced victim re-traumatisation and fewer people inadvertently viewing CSAM.[213] Identifying known CSAM can also prevent contact abuse. Evidence suggests that people who view CSAM are more likely to go on to commit other sexual offences against children, including contact offences. By reducing the availability of CSAM, hash matching technology thus can reduce the propensity of offenders to seek contact abuse, and can prevent potential perpetrators from viewing CSAM in the first place and going down an abusive pathway.[214] Moreover, by identifying known CSAM, services are also able to identify previously unknown CSAM, due to the proximity of known CSAM to unknown CSAM. In this way, hash-matching technology leads services to identify CSAM that was not previously known to law enforcement agencies and can help identify children and/or abusers, which can result in children safeguarded ('safeguards') and arrests.[215]

A14.23 Below, we attempt to quantify the number of children that could be safeguarded from contact CSA if a service implements hash matching. Drawing on existing estimates of the social cost of contact CSA, we then value the expected benefit from a reduction in contact CSA. No analogous estimates exist of other social costs of CSAM existing on a service, so we do not attempt to monetise other benefits associated with removing CSAM from a service. Instead, we limit our quantitative analysis to focus on the reduction in contact CSA.

A14.24 In paragraph A14.20, we explained how we estimated the expected number of CSAM reports that services would make if they implemented hash matching. We now estimate the expected number of safeguards that would result from the reports. What matters for the number of safeguards is the number of *actionable* reports a service makes.[216] As above, we can model the expected number of actionable reports by a service that implements hash matching based on the historical number of reports by services that already implements hash matching. Taking the same potentially high-risk service modelled above as our example, we calculate that this service made 0.0001 actionable reports per user in 2022.[217] As the Act constrains Ofcom from including measures in Codes that recommend the use of proactive technology to analyse user-generated content communicated privately, we assume that only 40% of these actionable reports stem from content communicated

[213] A survey commissioned by Ofcom found that 3% of adults and 5% of children said they had encountered CSAM online in the previous year. Source: Ofcom, 2020. Internet users' experience of potential online harms: summary of survey [accessed 09 September 2023].
[214] Protect Children (Insoll T., Ovaska A., Vaaranen-Valkonen N.), 2021. CSAM Users in the Dark Web: Protecting Children Through Prevention. [accessed 09 September 2023].
[215] Industry experts told us that this is a common channel through which services identify unknown CSAM.
[216] Reports are actionable based on a set of criteria used by NCMEC analysts to identify referrals where the reporting company has provided sufficient information to evidence a crime has been committed and require review and assessment by law enforcement agencies.
[217] This is based on the number of actionable CSAM reports submitted by the service to NCMEC that were subsequently passed on to the NCA, relative to the number of UK users.

publicly, in relation to which the proposed measure would apply.[218] This implies that, for a similarly high-risk service, we would expect 0.00004 actionable reports per user, when excluding reports based on private communications.

A14.25 To estimate the additional number of safeguards based on the number of actionable reports, we calculate the frequency with which actionable reports lead to safeguards. In 2022, the NCA received 50,721 actionable referrals from NCMEC.[219] In the same year, UK law enforcement protected or safeguarded 13,477 children against CSA[220], of which we assume that 40% depend on actionable reports received as a result of hash matching.[221] This implies that 0.1 children are safeguarded against contact CSA for every actionable report received by the NCA. If there are 0.1 safeguards per actionable reports, and if (as in the above example) there are 0.00004 actionable reports per user, then this implies that a child would be safeguarded for every 0.000004 users. This rate will vary by a service's risk level.

A14.26 We have drawn on estimates produced by the Home Office of the financial and non-financial (monetised) harm relating to all children who began to experience contact sexual abuse, or who continued to experience contact sexual abuse, in England and Wales in the year ending 31st March 2019. The total monetised harm per victim was estimated to be £89,000. Accounting for inflation, this figure would be £101,700 in 2022 prices.[222] This value allows us to monetise the benefit of reducing contact CSA. For example, if a service has 700,000 UK users and if (continuing with the above example) the service's risk level is such that it would make 0.00004 actionable reports per user, then we would expect the service to make 28 actionable referrals per year which, assuming 0.1 safeguards per actionable reports, we would expect to result in 3 children potentially safeguarded against contact CSA each year as a result of the service implementing the measure. This is equivalent to an annual benefit of £306,000.

[218] This assumption is based on NCMEC's estimate that "more than half of its CyberTipline reports will vanish with end-to-end encryption". For these purposes it is assumed that content communicated privately would be the content that is end-to-end encrypted. Source: NCMEC. End-to-end encryption: Ignoring abuse won't stop it. [accessed 25 September 2023].

[219] NCA provided us with data on reports passed to them from NCMEC. Source: [**CONFIDENTIAL ✂**].

[220] National Crime Agency, 2022. Annual Report and Accounts: 2021 – 2022. [accessed 05 June 2023].

[221] An industry expert [**CONFIDENTIAL ✂**] told us that 30-50% of national arrests and safeguards depend on actionable reports received as a result of hash matching.

[222] This is a conservative estimate of the social cost of contact CSA as it does not include long-term health impacts and does not apply to fatal CSA. Source: Home Office (Radakin, F., Scholes, A., Soloman, K., Thomas-Lacroix, C., Davies, A.), 2021. The economic and social cost of contact child sexual abuse. [accessed 05 June 2023]. The figure is broadly consistent with another study which estimated that the average lifetime cost for victims in the US of non-fatal CSA is $282,734 and $74,691 for female and male victims, respectively. Source: Letourneau E., Brown D., Fang X., Hassan A., Mercy J., 2018. The economic burden of child sexual abuse in the United States, *Child Abuse Neglect,* 79. [accessed 05 June 2023].

Table A14.5. Illustrative benefits of three hypothetical services implementing the measure

	File-storage and file-sharing service that reaches 70,000 UK users	High risk service that reaches 700,000 UK users	Medium-risk service that reaches 7,000,000 UK users
Value of expected reduction in contact CSA in the UK (annual)	£76,000	£306,000	£1,036,000
Other benefits	Non-monetised benefits include: • Fewer victims re-traumatised (e.g. due to image proliferation) • Fewer victims re-victimised (e.g. due to doxing and harassment) • Fewer offenders viewing CSAM and fewer people inadvertently viewing CSAM, which have been linked to contact abuse		

Source: Ofcom analysis, various sources

Additional supporting analysis

A14.27 Based on the assumptions set out above, we can illustrate the benefits and costs of a service implementing hash matching for CSAM given a service's user base and risk level. We present the illustrative benefits and costs for three hypothetical services, designed to align with the provisional conclusion in Chapter 14. In the below table, we compare costs and benefits after projecting them over a 10-year appraisal period and discounting by 3.5%.[223]

Table A14.6. Illustrative costs and benefits of three hypothetical services implementing the measure

		File-storage and file-sharing service that reaches 70,000 UK users	High risk service that reaches 700,000 UK users	Medium-risk service that reaches 7,000,000 UK users
Costs (present value)	Low-cost profile	£247,000	£453,000	£1,085,000

[223] For this analysis, we have not added financing costs (i.e. the cost of capital) for one-off build costs. This reflects that the one-off costs are comparable in magnitude to annual on-going costs and so only represent a small fraction of total costs. As such, this approach does not materially change any conclusions. When discounting, we have assumed that one-off costs are incurred in the first year and that on-going costs and benefits occur in subsequent years.

		File-storage and file-sharing service that reaches 70,000 UK users	High risk service that reaches 700,000 UK users	Medium-risk service that reaches 7,000,000 UK users
	High-cost profile	£1,118,000	£1,603,000	£2,177,000
Benefits (present value)		£558,000	£2,251,000	£7,615,000
Benefit cost ratio	Low-cost profile	2.3	5	7.0
	High-cost profile	0.5*	1.4	3.5

Source: Ofcom analysis, various sources

*We have presented the estimated costs for all hypothetical services based on both the low-cost and high-cost profiles, but we consider it unlikely that a service that reaches 70,000 users would have a high-cost profile and that a that reaches 7 million users would have a low-cost profile.

A14.28 The table shows that, even if we only consider the benefit of hash matching from reducing contact CSA, the benefits could exceed the direct costs for even the smallest services that we are provisionally recommending hash matching to.

A14.29 While the costs are not insignificant, the quantitative evidence supports our existing position that because the harm is so significant, the measure is proportionate for many sizes and kinds of service, including some services with a small user base. For example, for a hypothetical high-risk service with 700,000 UK users, we estimate hash matching could lead to 28 actionable reports per year leading to the safeguarding of 3 children per year. In monetary terms, this translates into benefits of £2,251,000 when projected over 10 years. This exceeds the estimated costs, even if the service has a high-cost profile. The only scenario presented above in which the monetised benefits do not exceed costs is for file-storage and file-sharing services with 70,000 UK users and a high-cost profile, but we consider it unlikely that a service this small would have a high-cost profile. In general, the analysis illustrates that the benefits of the measure are likely to be far greater than the costs, including for services with small user bases where the risk of harm is great.

A14.30 We recognise that there are limitations to this type of analysis and there will be benefits beyond what we have been able to account for quantitatively. However, this analysis supports our provisional conclusion by demonstrating that even when a limited range of benefits are considered, the costs are justified even for the smallest services that we are provisionally recommending hash matching to.

Further analysis on the options for applying measures that reduce the risk of grooming

A14.31 Chapter 18 on U2U default settings and users support for child users recommends several measures which focus on reducing the risk of grooming for some U2U services. In paragraph 18.89, we consider three possible options for how we could target the default setting measures:

a) Option 1: Apply the measures to all large services which have a high or medium risk of grooming.
b) Option 2: Apply the measures to: (i) all services which have a high risk of grooming AND at least 25,000 child users; and (ii) all large services which have a medium risk of grooming.
c) Option 3: Apply the measures to: (i) all services which have a high risk of grooming and (ii) all large services which have a medium risk of grooming.

A14.32 To help evaluate which of these is appropriate, below we consider available information on the different costs and benefits associated with each of these options. All the options apply the measure in the same way for medium risk services, so we have focused on the costs and benefits of applying the measure to high risk services.

A14.33 For the reasons set out below, we are confident that Option 1 would be a proportionate intervention, but we consider it would not go far enough. The choice between Options 2 and 3 is more evenly balanced, but our provisional view is that we should adopt Option 3.

Estimating the costs and benefits for different options

A14.34 As one input to assessing the options, we have put monetary values on some of the likely benefits and costs of each of the options, while recognising that this type of analysis has significant limitations. In our assessment below, we have made various assumptions that are often based on limited and uncertain information. We account for uncertainties in part by considering ranges for specific inputs and by developing different scenarios.

A14.35 As is the case for our proposed measure for CSAM hash matching, estimating and quantifying the economic and social cost of CSEA offences, and the likely benefits of reducing grooming, is challenging. As described in the CSEA chapter of the Register of Risks, 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 sexually abuse other children or to engage in sexual acts, including penetrative acts. Individuals involved are often forced to share imagery of the abuse with the perpetrator as ‘first-generation’ CSAM.

A14.36 We recognise that the impacts of grooming on each victim and survivor are different and trying to put a monetary value on this is a vast simplification. The only estimate we are aware of that can be related to an aspect of grooming is an estimate by the Home Office of the economic and social cost of contact CSA. This is an estimate of £101,700. Unsurprisingly, the largest component of this estimate is the physical and emotional harm caused to the victim. However, this is a conservative estimate as it does not reflect the harm where CSA

has led to death and does not fully account for the long-term mental health impacts on victims.[224]

A14.37 Because we are only aware of a quantified estimate for contact CSA, we have only been able to quantify this aspect of the many harms that can arise from grooming, and hence the benefit of any reduction in this harm. This means our estimate of the benefits from reducing grooming will be materially understated.

A14.38 The impact of grooming on victims and survivors is complex and the experience of harm and cost to an individual is a personal one that will vary depending upon may different factors including, but not limited to, personal characteristics, coping mechanisms and the type of harm they have been exposed to. It is recognised that victims and survivors who have experienced similar CSEA harms may describe the effect to them differently, from having minimal impact to having a significant impact on them. For some, the potential impact of grooming can be severe and lifelong, regardless of whether it leads to contact sexual abuse offences. In some cases, online grooming can also lead children to self-harm and taking their own life.[225] We have not aimed to include the full spectrum of harms that can arise from grooming. This is due to the complexities of estimating both the social cost of different types of harm that are a result of grooming, and also the number of instances of each type of harm which occur each year.

Estimating benefits from reducing contact CSA from grooming

A14.39 Our approach estimates a lower bound of the benefits to society of an individual service implementing the default setting measure by combining:

a) The expected number of contact CSA instances resulting from grooming on a service;
b) The expected reduction in such instances resulting from the measure; and
c) The social cost of contact CSA.

Equation 1: Expected annual benefits from the default setting measure in reducing contact CSA from grooming[226]

*Annual benefits (from reducing contact CSA from grooming) = Expected number of contact CSA occurrences resulting from grooming on a service * expected reduction in such occurrences resulting from the measures * social cost of one contact CSA occurrence*

[224] The original study estimated was £89,240, in 2018/19 prices and we have increased to put in 2022 prices. Home Office (Radakin, F., Scholes, A., Soloman, K., Thomas-Lacroix, C., Davies, A.), 2021. *The economic and social cost of contact child sexual abuse.* The figure is broadly consistent with another study which estimated that the average lifetime cost for victims in the US of non-fatal CSA is $282,734 and $74,691 for female and male victims, respectively. Letourneau et al., 2018. *The economic burden of child sexual abuse in the United States.*

[225] Example of such deaths have been reported in the press. For example, Carrell, S. 2013. Scotland police investigate 'online blackmail' death of Fife teenager, *The Guardian,* 16 August. [accessed 20 September 2023]; Campbell, J. & Kravarik, J., 2022. A 17-year-old boy died by suicide hours after being scammed. The FBI says it's part of a troubling increase in 'sextortion' cases. *CNN*, May 23.; Yousif, N., 2022. Amanda Todd: Dutchman sentenced for fatal cyber-stalking, *BBC News,* 15 October. [All accessed 04 September 2023]

[226] Although we are only applying this equation to grooming that leads to contact CSA, a similar equation could be used for different types of grooming (eg, non-contact CSA, grooming resulting in fatalities). The combined benefits that result from all equations would then provide an estimate of the total benefits from the measures. As mentioned above, we have concentrated our quantitative analysis on grooming resulting in contact CSA because we are not able to estimate input assumptions for any other types of harm from

A14.40 We have estimated the expected number of grooming occurrences on an online service that lead to contact CSA by considering: the number of child users on that service; the percentage of those children who are subject to grooming leading to contact CSA each year; and how many separate services they are likely to visit over the course of a year.

A14.41 As not all children will be using the internet or using sites where grooming may occur (eg, very young children are less likely to use social media), we have estimated the number of children using services where grooming may happen. We refer to children who are using services where grooming may happen as 'at risk' children in Equation 2. We explain further below how we have estimated the number of 'at risk' children for the purposes of this analysis.

Equation 2: Expected number of grooming occurrences that lead to contact CSA on an online service

*Expected number of grooming occurrences on an online service that lead to contact CSA = Number of child users on the individual service * (% of 'at risk' UK children that experience grooming leading to contact CSA annually/Average number of services visited by UK Children annually where there is a risk of grooming)*

A14.42 To calculate the above equations, we combine our estimate of the social cost of contact CSA with estimates of the following:

i) Percentage of 'at risk' children that experience grooming leading to contact CSA in a year.[227]
ii) Number of child users on the individual service.
iii) Average number of sites a child visits in a year where there is a risk of grooming.
iv) Expected reduction in grooming incidents due to safety measures.

A14.43 We outline below how we have estimated each of these inputs. We use these inputs to estimate the potential benefits of the options we consider, alongside an estimate of the costs of implementing the measures. Given the uncertainty around the exact magnitude of our inputs, we tested the costs and benefits across several scenarios.

Percentage of 'at risk' children that suffer contact CSA as a result of grooming annually

A14.44 We define 'at risk' children for the purposes of this analysis as those who use the internet and visit sites where we expect that there may be a risk of grooming. We first estimated the total number of 'at risk children in the UK and then considered more specifically the percentage of those children likely to experience grooming leading to contact CSA each year.

Number of children at risk of grooming

A14.45 The Office of National Statistics forecast there would be 12.2 million children aged 3 to 17 in the UK in 2023.[228]

---

grooming. Despite this, we have qualitatively considered the potential benefits from other types of harm from grooming when concluding on the proportionality of this measure.

[227] We define 'at risk' children for the purposes of this analysis as those at risk of grooming because they use online services where grooming can take place.

[228] Population projections published in January 2022 by the Office of National Statistics. [accessed 28 September 2023].

A14.46 Of these children, we only assume that they are susceptible to grooming once they start using social media and other services. We estimated the percentage of children in each age category that have used social media, from Ofcom's 2023 Children's media use and attitude report.[229]

A14.47 This gives us an estimate of 7.8m children who may be at risk of grooming.

Percentage of children who suffer contact CSA as a result of grooming annually

A14.48 The UK government has estimated there are approximately 113,000 cases of contact CSA per year in England and Wales.[230] This estimate is based on a survey of young people and their care givers, which asked about their experiences of sexual abuse in the year prior to the survey.[231] We increased this number to make it applicable to be used as a UK wide estimate[232] and then combined it with an estimate used by DCMS in the impact assessment for the Online Safety Bill that 20% of all recorded contact CSA offences had an online element.[233] We assume that all cases with an online element would have involved grooming. Combining these assumptions indicates there are approximately 25,000 cases of contact CSA each year that are likely to have an online element, and are therefore indicative of potential online grooming. This is equivalent to assuming that approximately **0.3%** of the 7.8m children who could be at risk of grooming in the UK are likely to experience contact CSA as a result of grooming each year. The proportion who experience contact CSA at some point during their childhood would be higher than this per year percentage.

A14.49 We recognise that this assumption is conservative and is likely to understate the true prevalence of contact CSA resulting from grooming in the UK. In particular, the annual number of CSA offences was based on a survey and is likely to understate CSA[234] and the proportion of recorded CSA cases that had an online element may be greater than 20%.[235]

---

[229] 30% of 5-7 year olds, 63% of 8-11 year olds, 93% of 12-15 year olds and 97% of 16-17 year olds use social media apps or sites and therefore may be at risk of grooming. Source: Ofcom, 2023. *Children and Parents: Media Use and Attitudes, pages 21-31.* [accessed 28 September 2023].

[230] Home Office, Dec 2021. *The economic and social cost of contact child sexual abuse*, Section 3.2 [accessed 28 September 2023].

[231] The estimate of contact CSA occurrences is ultimately derived from Radford, L., Corral, S., Bradley, C. and Fisher, H., 2013. The prevalence and impact of child maltreatment and other types of victimization in the UK [accessed 28 September 2023]. This involved a survey of 6,196 children, young people and caregivers that was conducted in 2009. The victimisation rate from that survey was then applied to the child population in England and Wales in 2018 to estimate the number of occurrences.

[232] We increased this based on the number of estimated number of children across the UK compared to England and Wales. This resulted in an estimate of approximately 126,000 cases of contact CSA per year across the UK.

[233] The percentage of contact CSA cases that have an online element was estimated using data recorded by the police over the period April 2020 to March 2021. DSIT, 2022. Online Safety Bill impact assessment, paragraph 267 [accessed 28 September 2023].

[234] For children aged under 11, the survey was answered by caregivers. Therefore CSA occurrences are unlikely to be included if the caregiver was unaware of the abuse or was the abuser themselves. Source: Home Office, Dec 2021. *The economic and social cost of contact child sexual abuse*, Section 3.1 [accessed 28 September 2023].

[235] The true level of online offences may be higher than 20%, due to issues with recording whether an offence had an online element and the increasing proliferation of online technology. DSIT, 2022. Online Safety Bill impact assessment, paragraph 268 [accessed 28 September 2023].

Number of online services a child visits where there is a risk of grooming

A14.50 A child who suffers contact CSA as a result of grooming in a year is likely to have visited several online services over the course of that year. Therefore, we need to understand how many services a child is likely to visit over the year, where there could be a risk of grooming, to be able to estimate the risk of grooming on an individual service.

A14.51 Ipsos iris reported on several thousand [**CONFIDENTIAL ✂**] organisations that were visited by 15-17 year olds in December 2022. This estimate is likely to be representative of the vast majority of expected internet sites and apps visited by children of that age category. Note that there are only less than 100 ([**CONFIDENTIAL ✂**]) 15-17 year olds in a sample of monthly data, so these results are indicative only and intended to illustrate the real world rather than intending to offer an accurate representation of user behaviour.[236]

A14.52 To estimate the number of online services a single child is likely to have visited we used unweighted raw data from 15-17 year olds and calculated the average number of organisations or platforms visited across the sample.

A14.53 We removed some online services from the analysis when we considered there was limited risk of grooming. We did this in two ways:

a) We removed online services where we categorised the organisation as not likely to have a risk of grooming based on their business category provided by Ipsos Iris.[237]
b) We removed online services where the average time spent by a 15-17 year old over the course of the month was less than 10 minutes. For these platforms, we made a decision that they were unlikely to have significant risk of grooming due to the short time available for any interaction.

A14.54 The results of this analysis suggests that the average number of individual organisations that a 15-17 year old visited over the course of a month was 10.

A14.55 We consider this value provides a useful indicator of the number of online services which a child at risk of grooming is likely to visit. However, we recognise that:

a) The analysis is based on data from surveying a limited number of 15-17 year olds. There is significant uncertainty in extrapolating results from a survey with a small sample size to be representative of the whole 15-17 population. In addition, the analysis only covers 15-17 year olds, which means that it is not reflective of the number of sites that younger children would typically visit, which we expect would be lower.[238]

---

[236] 'Organisations' are the parent companies of the groups of websites and apps; for instance, Alphabet organisations include Google Search, Gmail and YouTube; Meta includes Facebook and Messenger, Instagram and WhatsApp. We have undertaken our analysis at an organisation level but recognise that this could lead to an underestimate of the number of individual online services visited. However, we also consider that there could be cost synergies when implementing the measure across an organisation, which provides a further reason why we consider it is reasonable to consider this at an organisation level.

[237] Ipsos Iris provides classifies all online services into different categories. Using a manual process we estimated whether these categories were likely to indicate a site at risk of grooming (eg, 'Social media', 'gaming', etc.) or limited risk of grooming (eg, 'Weather', 'Estate agents', etc.).

[238] Our Media Use and Attitudes report suggest that "…16-17-year-olds are branching out in media, using a wider and more diverse diet of apps and sites." Source: Ofcom, 2023. *Children and Parents: Media Use and Attitudes, pages 28.*

b) The analysis only covers the viewing habits of 15-17 year olds over the course of one month. Although we expect that generally child users will visit the same websites from month to month, there may also be some additional services they visit on specific months. Including this effect would tend to increase the estimate of the number of sites visited.
c) As described above in paragraph [A14.51] and footnote [40] our analysis only considers the number of 'organisations' which are visited, not individual online services themselves. Including this effect would tend to slightly increase the estimate of the number of sites visited.

A14.56 We consider the strongest caveat of those listed above is that the estimate could be overstating the number of services that younger children visit over a year, where they are likely to be at risk of grooming. However, using a lower estimate for the number of sites would increase the value of the benefits we would expect from the measure for each service. This means that the estimate of 10 services can be considered a relatively conservative assumption, in the sense that the real world number is likely to be lower which would tend to increase the estimates of the benefits of the measures.

A14.57 After considering the results of the analysis and the various factors outlined above, we think that a reasonable assumption for the purpose of our analysis is that each child is likely to visit approximately 10 online services where there is a risk of grooming.

Expected reduction in grooming due to implementation of the measures

A14.58 We believe that the suggested measures will be effective at reducing grooming by introducing friction into the grooming process. However, they are unlikely to fully eliminate the incidence of grooming:

a) The measures are focused on default settings that can be turned off.
b) The measures will only apply to services which can identify children on their platform.
c) There are likely to be other pathways for perpetrators to identify children and start a grooming process.

A14.59 For these reasons and because the actual effectiveness of the measure is particularly uncertain, we assume that in combination all of the elements of the default setting measure could reduce the extent of contact CSA as a result of grooming by between 5% and 10%.[239]

A14.60 Based on these assumptions, we have developed a high benefits scenario, where the measures are assumed to reduce contact CSA from grooming occurrences by **10%** and a low benefits scenario where the measures reduce it by **5%.**

Estimating the benefits of reduced contact CSA for each option

A14.61 We used the assumptions outlined above to estimate the benefits from reduced contact CSA from grooming for each of the options we are considering. As the options apply the measures equally to medium risk services, we have focused on the impact on high risk services, where the difference between the options is as follows:

a) Option 1: Apply the measures to all large services which have a high risk of grooming.

[239] We assume services apply all elements of the default setting measure outlined in the Chapter 18.

b) Option 2: Apply the measures to all services which have a high risk of grooming AND at least 25,000 child users.
c) Option 3: Apply the measures to all services which have a high risk of grooming.

A14.62 Option 1 would only apply the measures to large services. We propose across the codes to define a 'large' service as one with more than 7 million UK users.[240] To estimate the benefits from the measure we needed to estimate the number of children on the service. We assume that a large service would have at least 1 million child users aged 3-17 based on an assumption that around 15% of UK internet service users are aged 3-17.[241] This suggests large services are likely to have at least **1 million child users**. This is about 8% of all 3-17 year olds in the UK.

A14.63 Option 2 would only apply the measures to high risk services with more than 25,000 child users. This is approximately 0.2% of all 3-17 year olds in the UK.

A14.64 Option 3 would apply the measures to all high risk services, regardless of the number of child users.

A14.65 To test these options, we estimated the benefits and costs for services who had 1 million child users and 25,000 child users. We used these estimates together with a qualitative assessment of costs and benefits for services with less than 25,000 users as part of our assessment of the proportionality of the three options.

A14.66 We considered the impact on different sizes of services sequentially. Ie, first considering services with 1 million child users, then considering services with 25,000 child users, and then for all services. Stepping through these in turn has informed our consideration of whether to propose Option 1, Option 2 or Option 3.

A14.67 We present in this section the results of our analysis, including:

a) A summary of the assumptions in the quantitative analysis used to estimate benefits.
b) A summary of the assumptions in the quantitative analysis used to estimate costs.
c) The results of our quantitative analysis assessing a service with 1 million child users.
d) The results of our quantitative analysis assessing a service with 25,000 child users.
e) Our overall provisional conclusion on which option to propose in the Codes.

Summary of input assumptions used to estimate benefits

A14.68 Table A14.7 below summarises the assumptions we used to estimate the benefits for a high and low scenario.

[240] See from paragraph 11.51.
[241] We have estimated this number to be 15% using ONS population data by age category and ONS data on internet usage by age.

Table A14.7: Estimate of benefits from reducing contact CSA from reduced grooming on a service

	Benefits	
	Low benefit scenario	**High benefit scenario**
Harm that arises from a contact CSA occurrence	£101,700	£101,700
Percentage of children that experience contact CSA as a result of grooming in a year	0.3%	0.3%
Average number of services that a child visits each a year where they are at risk of grooming	10	10
Expected reduction in grooming occurrences due to default setting measures	5%	10%

Source: Ofcom analysis

Summary of input assumptions used to estimate costs

A14.69 We outline in Table 18.1 in Chapter 18 our assumptions for the potential range of costs that could arise from implementing the default setting measure. This suggests a potential range for the one-off costs between £10,000 and £300,000.[242]

A14.70 We applied the following three cost assumptions in the analysis:

a) **Services with 1 million child users**: We assumed that costs are at the top of the range (ie £300,000 one-off cost).
b) **Services with 25,000 child users (low estimate):** We assumed that costs are at the bottom of the range (ie, £10,000).
c) **Services with 25,000 child users (high estimate):** We assumed that costs are £50,000.

A14.71 We have included two cost scenarios for services with 25,000 child users because although generally we would expect smaller services to be towards the lower end of the range and larger services to be towards the top, there are also several non-size related factors which would affect the level of costs that might apply to different services.[243]

A14.72 Therefore, there might be some very small services who would have costs higher than the very bottom of the range, but for which the top of the range is also unlikely to be appropriate. This is because the upper end of our cost range includes significant overhead and coordination costs[244] that are likely to be much lower for smaller services. Therefore, we have included an additional 'high estimate' for services with 25,000 child users where we assume the one-off cost is £50,000.

[242] Note that for the quantitative analysis we do not estimate any indirect costs.
[243] For example, the complexity of their systems or the extent to which they can easily switch default settings for individual users.
[244] For more explanation on overhead and coordination costs, please see Chapter 18, paragraphs 18.54-18.55.

A14.73 As well as the one-off costs we also include ongoing costs[245] associated with implementing the measures. The one-off cost will be incurred once as the measure is implemented, while the ongoing costs and benefits that arise from implementing the measures will occur annually. To ensure that costs can be considered alongside annual benefits, we estimated an annualised cost by calculating the total cost of the measures over a 10-year period, including the financial costs that we expect to be incurred for one-off investments,[246] and then converted to a flat annuitized rate to estimate the annualised costs of the investment.[247]

A14.74 We do not include any indirect costs within our analysis as we are unable to appropriately quantify them. However, we consider indirect costs (for both services and users) will be lower for smaller services, as they will tend to scale with the number of users. As we describe below, we think the case for recommending the measure for large services (option 1) is very strong, and we are confident it is proportionate to apply to such services even though we have not quantified the indirect costs. When considering the case for recommending for smaller services (as with option 2 and 3), these indirect costs will be much smaller and so not including them is less important.

A14.75 Table A14.8 outlines the cost assumptions we use in different scenarios including the estimate of costs on an annual basis.

Table A14.8: Estimated cost of default setting measures under different scenarios

	Costs of default setting measures, per service		
Assumptions			
	Low cost scenario	**High cost (smaller services) scenario**	**High cost (large services) scenario**
Initial cost of default setting measures	£10,000	£50,000	£300,000
Ongoing cost of default setting measures	£2,500	£12,500	£75,000
Annualised cost of the measures	**£4,000**	**£20,000**	**£118,000**

Source: Ofcom analysis

Quantitative assessment of services with 1 million child users (Option 1)

A14.76 We estimated the annual benefits for a service with 1 million child users by inserting the assumptions in Table A14.7 into equations 1 and 2. We used this together with the

[245] We assume that annual ongoing costs are equal to 25% of the one-off costs. For an explanation, please see paragraph A14.12.

[246] We have assumed a real pre-tax financial cost of capital of 7% for these companies, but note our findings presented below are not very sensitive to the assumed cost of capital. A WACC of 7% is broadly consistent with an estimate the CMA used for Google and Facebook as part of their 2020 Market Study on online platforms and digital advertising. *2020 Online Platforms and Digital Advertising market study final report, Appendix D*, pages D39-D41 [accessed 15 September 2023].

[247] We annualise the costs by assuming that the one-off costs are akin to an investment that is paid back at a consistent rate in real terms over 10 years. This payment includes both the financing costs and repayment of the initial investment cost. We also assume that the annual benefits and ongoing costs would start in the same year that the one-off investment takes place. This is because we expect that the default setting measures could be implemented and become effective over a relatively short timescale.

annualised cost to estimate illustrative benefit cost ratios for a service of this size when applying the measure.

A14.77 The estimated benefit cost ratios for both the high and low benefit scenarios for a service with 1,000,000 child users are given in Table A14.9.

Table A14.8: Indicative benefit cost ratios for services with1,000,000 child users

	Annual Estimated benefits	Annualised cost	Indicative benefit cost ratio
High benefits scenario	£3,246,000	£118,000	28
Low benefits scenario	£1,623,000	£118,000	14

Source: Ofcom analysis

A14.78 The benefit cost ratios in Table A14.9 show benefits much greater than costs for large services with above 1 million child users, even though our estimate of benefits only considers the benefits that would arise from a reduction in contact CSA that is a result of online grooming. As there are other benefits, the total benefits from the measure for a service of this size is likely to be significantly higher.

A14.79 We also assumed that the expected costs were at the very top of our estimated range. This conservative approach combined with the high ratio of benefits to costs gives us significant confidence that applying the measures to all large services which have a high risk of grooming is likely to deliver significant benefits, and that Option 1 would be proportionate. Additionally, we also considered our estimates above in the case of whether to apply the measure to large medium risk services. Although the analysis does not specifically cover medium risk services, our analysis for large high risk services shows very high estimated benefits in comparison with cost. This gives us confidence that it is also likely to be proportionate for medium risk services, even though the benefit from the measure is likely to be slightly lower for medium risk services compared to high risk services. Additionally, the high benefit cost ratios indicate that applying the measure to large medium risk services is also likely to be proportionate, even though the benefit from the measure is likely to be slightly lower for medium risk services compared to high risk services.[248]

A14.80 However, the significant level of benefits relative to cost indicates that choosing Option 1 is unlikely to go far enough, and the measure is also likely to be proportionate at much lower levels of child users. In particular:

a) As set out in the Register of Risks, grooming is not confined to the largest platforms. Perpetrators can target any platform where there are children, regardless of size. Option 1 would therefore leave a material part of the problem of grooming on high risk services unaddressed.
b) Moreover, it would likely have displacement effects. If the largest services take steps to improve protections against grooming, this may result in perpetrators shifting to focus on targeting children on smaller services. As we show in the Register of Risks, we have observed displacement effects of this nature occur when large services have moved to improve protections against other harms.

[248] We expect the benefit to be slightly lower for medium risk services because we expect that they would have a slightly lower prevalence of grooming, that leads to contact CSA, than we have used in the analysis.

A14.81 Therefore, our provisional view of Option 1 is that to apply the measure to large services only is unlikely to go far enough.

Quantitative assessment of services with 25,000 child users (Option 2)

A14.82 Using the same method as outlined above, we estimated the annual benefits for a service with 25,000 child users and then combined this with the annualised cost to estimate indicative benefit cost ratios.

A14.83 The estimated benefit cost ratios for the relevant four scenarios is given in Table A14.10.

Table A14.10: Indicative benefit cost ratios for services with 25,000 child users

	Annual Estimated benefits	Annualised cost	Indicative benefit cost ratio
High benefit/low cost scenario	£81,000	£4,000	21
Low benefit/low cost scenario	£41,000	£4,000	10
High benefit/high cost scenario	£81,000	£20,000	4
Low benefit/high cost scenario	£41,000	£20,000	2

Source: Ofcom analysis

A14.84 Table A14.10 shows that the estimated benefits are greater than costs across all four scenarios we have assessed for services with 25,000 child users. One scenario (low benefit/high cost) shows benefits to be only moderately higher than costs, however as the benefits are likely to be significantly understated, we consider these results indicate that, at the very least, the measure is likely to be proportionate for services with 25,000 child users or more.

A14.85 Overall, we consider that the results of the quantitative analysis illustrate that the measures are likely to result in benefits that are significantly in excess of the costs for Option 2. However, what the quantitative analysis does not do is indicate whether the measures are also likely to be beneficial for services which have even lower numbers of child users, as it is unable to capture all of the factors that impact whether the measure is proportionate for those smaller services. We consider that question in the following section.

Qualitative assessment of Option 3 and provisional conclusion

A14.86 The choice between applying the measure to services with more than 25,000 users (Option 2) or to extend further and apply it to all high risk services (Option 3) is more evenly balanced. Option 2 would not apply the measure to services with very few children on them. This would reduce the risk of inadvertently imposing disproportionate costs on services where grooming did not in practice occur frequently. Option 3 would provide more comprehensive coverage against grooming, as it would capture all services which have

[249] Note the values in this column may not precisely match with the presented benefit and costs in the other columns due to rounding.

identified as having a high risk of grooming and reduces the potential for harm due to displacement of perpetrators to very small services.

A14.87 A key factor in our proposals is that the benefits estimated above for services with 25,000 child users are likely to be significantly understated. In particular:

a) We have only estimated the costs of contact CSA offences resulting from grooming, which we acknowledge only represents a small proportion of all grooming occurrences. We outline in Chapter 18 a US study that found that 17% of participants experienced sexual solicitation as a youth 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.[250] Our analysis does not include any other harms that could result from the many other instances of grooming which do not result in contact CSA.[251]
b) The estimated harm of individual contact CSA offences resulting from grooming as set out in this annex are likely to be understated. This is partly because we consider the survey from which the number of contact CSA incidents has been derived is likely to understate the amount of contact CSA.[252] Additionally the estimate of the social cost of one instance of contact CSA understates the full impact on an individual, as it only reflects the cost of non-fatal CSA and does not account for long-term mental health costs or other social and emotional impacts.

A14.88 On balance, our provisional view is that we should adopt Option 3. This has been informed by both the quantitative analysis and a wider qualitative assessment of the factors that affect the proportionality of the measure. The reasons for applying the measure to all high risk services are:

a) The widespread nature of the threat grooming poses and the severity of the harms it can lead to. Child sexual abuse is a horrific crime which can have a severe and lifelong impact. This argues for applying Option 3 rather than Option 2, not least given that the impact of grooming is so material that the measure would only need to prevent a very small number of cases of grooming on any given service for the benefits to justify the costs of the measure.
b) As described above, it is likely that the true benefit would be higher than we have been able to estimate, and potentially significantly higher than we have modelled for a service with 25,000 child users. This indicates it would be proportionate to apply the measure to services with potentially much lower numbers of child users.
c) As set out in the Register of Risks, perpetrators target services of all sizes where there are children, even very small services.[253] Option 2 could therefore still leave important gaps in protection. A particular risk is the potential for perpetrators to move to using

[250] Please see Chapter 18, paragraph 18.3.
[251] We recognise that the harms that result from these incidents are not all likely to be as material as the harm we have quantified from contact CSA, however as with all CSEA offences the impact of grooming can be severe and lifelong, regardless of whether or not it leads to contact sexual abuse offences. In some cases, the harm may also be more material as online grooming can lead children to self-harm and taking their own life. Example of such deaths have been reported in the press. For example, Carrell, S. 2013. Scotland police investigate 'online blackmail' death of Fife teenager, *The Guardian,* 16 August. [accessed 20 September 2023]; Campbell, J. & Kravarik, J., 2022. A 17-year-old boy died by suicide hours after being scammed. The FBI says it's part of a troubling increase in 'sextortion' cases. *CNN*, May 23.; Yousif, N., 2022. Amanda Todd: Dutchman sentenced for fatal cyber-stalking, *BBC News,* 15 October. [All accessed 04 September 2023]
[252] See paragraph A14.49 above.
[253] Please see, Volume 2: Chapter 6(CSEA) paragraphs 6C.34 to 6C45.

smaller services if it were easier to connect with children on such services because they were excluded from the measure. This risk is not captured by our quantitative analysis above.

d) Option 3 only targets the measure at smaller platforms if they are at high risk of grooming. Given the severity of the harm, where a service is genuinely high risk there is a strong argument that it should not be exempt from providing children with protection, regardless of its size. The fact that the option places the most onerous obligations on the highest risk services is an important factor in our assessment of proportionality.
e) Relatedly, we also consider that some smaller services that are high risk could be particularly unsafe to child users on that platform (ie, they have a higher risk of grooming than the average service we have included within our analysis). For those services, the proportion of child users who are targeted by perpetrators may be higher than the cross-sector averages we have calculated above. Consequently the potential benefits (per child user) from introducing the measures are likely to be higher than is implied from the analysis. This further indicates that applying the measure to all services is likely to be appropriate to ensure we capture these types of services within the measures.
f) The costs will tend to be towards the lower end of the range we estimated for smaller businesses because such businesses will not have material high overheads and coordination costs associated with implementing the measure. This strengthens the argument that our proposal is proportionate. For services which are not large (Ie, those with a total user reach of less than 7 million) to be captured by our measures, they would need to both be able to identify child users and have assessed themselves as 'High risk' for grooming in their risk assessment.[254]
g) It is likely that smaller services, and particularly if they are services with more risky network expansion and connection functions, are likely to be at a relatively early stage of development and in the process of growing their user base. We consider that there could be some benefit from providing certainty on the required safety measures when high risk functionalities are incorporated into online services, whatever the number of child users. This approach could be beneficial if it means services are not required to update systems and interfaces once they pass a certain number of child users. We also consider that for a new, growing service, the additional cost that it incurs from making a greater upfront investment due to the measures, compared to the costs which it would incur once it passes a certain size, is likely to be small.

[254] Assessing as high risk for grooming includes having direct messaging functionalities alongside at least one other high risk factor, like network expansion prompts, connection lists or evidence of existing systematic grooming. Please further information on this please see Chapter 18, paragraph 18.85.

A15. Automated Content Moderation (U2U): design of measures

A15.1 This annex discusses the design of the measures proposed in Chapter 14 on Automated Content Moderation for U2U services for hash matching for CSAM, URL detection for CSAM URLs, and keyword detection regarding articles for use in frauds.

Hash matching for CSAM

A15.2 This section discusses each of the following areas in turn:

- The type of hash matching technology;
- The hash database used by services;
- The breadth of content that is scanned (and when) on the service (i.e., for new content or for all existing content);
- What provision should be made about the technical performance of the technology; and
- The use of human review in relation to content identified by the hash matching process

A15.3 The outline measure proposed based on these discussions can be found in Paragraph 14.42 of Chapter 14. A copy of the draft measure that we are proposing to include in our CSEA Code of Practice can also be found in Annex 7 to this consultation.

The type of hash matching technology

Cryptographic hash matching

A15.4 Cryptographic hash matching is highly accurate in detecting exact matches between two identical pieces of CSAM. It is highly unlikely that cryptographic hash matching would return any false positives, and any incorrect images that are surfaced through cryptographic hash matching are likely to be from images being incorrectly added to the hash database. As a result, if a service deployed cryptographic hash matching for CSAM alongside a high-quality hash database, and removed any content returned, it is highly unlikely that the service would incorrectly remove any user-generated content which is not CSAM.

A15.5 However, cryptographic hash matching is not able to detect images that are visually similar, or appear identical to the human eye, but which are not digitally identical. This may result in illegal content not being detected.

Perceptual hash matching

A15.6 By contrast, perceptual hash matching aims to create hashes that are very similar to each other for content that appears visually similar. This enables an estimation of similarity between the hashed contents based on the similarity of their hashes. Thus, perceptual hash matching may return an image if its hash is calculated as being similar to the hash of known CSAM in the hash database, but not identical. In practice, this means that perceptual hash

matching is more likely to detect more CSAM than cryptographic hash matching, as it can detect CSAM even where two hashes are not identical.

A15.7 However, as perceptual hash matching requires content to be similar, but not identical, this can cause false positive returns where the hashes of two items of content are similar, despite the contents themselves being dissimilar. This presents a higher risk that content which is not CSAM is incorrectly removed by services. To mitigate this risk, human review is commonly deployed to assess the accuracy of matches returned by perceptual hash matching. In addition, services can gather performance data and run dedicated tests to adjust the parameters of their hash function to test its accuracy and alter the false positive/negative rate appropriately.

A15.8 Perceptual hash matching may be more easily deployed by services with higher human review capacity, as they are more likely to have capacity to review potentially greater number of positive matches generated by perceptual hash matching, to determine the appropriate course of action for content that is returned by the system. Large services with more resource to invest in this technology may also use additional safeguards, such as using perceptual hash matching in tandem with machine learning-supported automated content classification systems to proactively detect and action such false positives.

A15.9 Furthermore, while some false positive content may be reported by the service to a reporting body at the point it is matched, there are additional assessment and review processes by those external bodies through the duration of the content escalation and investigation processes, including by NGOs and law enforcement, at which point false positive content is identified and, if appropriate, no further action is taken. While this does not preclude an impact on users' rights to privacy and freedom of expression, it does limit subsequent impacts on users' rights except where in accordance with the law.

Initial view

A15.10 We consider that perceptual hash matching is more effective at detecting a wider range of CSAM content than cryptographic hash matching, as it can result in far fewer false negatives. In addition, we consider that perceptual hash matching is more difficult to evade than cryptographic hash matching, which can be easily evaded. As a result, the use of perceptual hash matching is likely to result in more CSAM being detected.

A15.11 Moreover, the most popular proprietary solutions (e.g., Microsoft PhotoDNA, Google CSAI match) use perceptual hash matching technology, emphasising industry's preference for perceptual hashing. Perceptual hashing is widely supported across NGOs and industry and appears to be more widely deployed in practice.

A15.12 We considered whether enabling services to choose either or both technologies could offer some additional flexibility. However, the material difference in the effectiveness in detecting CSAM between cryptographic and perceptual hash matching, and the severity of the harm presented by CSAM, suggests that perceptual hashing may be the better way to address this harm.

The hash database

A15.13 The hash database that underlies any hash matching system is critically important to the overall functioning of the hash matching technology. The effectiveness and accuracy of hash matching technology are intrinsically linked to the scale and accuracy of the underlying database.

A15.14 We have therefore considered what provision should be included with regards to the hash database in any hash matching recommendation.

A15.15 We are aware that services already using hash matching take a variety of different approaches, for example:

- by acquiring hash databases from third-party providers, which in some cases may also perform the hash matching process on the service's behalf;
- by maintaining an internal hash database of hashes of content previously detected on their service;
- by layering various approaches to hash database sourcing and implementation to provide a more layered and robust response to CSAM detection.

A15.16 We are aware of a number of accessible options for services seeking to implement hash matching technology or to ingest hashes from external sources, such as from national and international NGOs.

A15.17 We are aware that the approach that services currently take depends on a number of factors, including:

- their ability to acquire access to external databases;
- their technical capacity to implement the technology; and
- the resource they have available to dedicate to all aspects of the hash matching process (including content review, takedown and action against users, and external reporting).

A15.18 For example, larger services with greater capacity and experience in running these systems may operate their systems entirely in-house, whereas a smaller service may outsource (parts of) the process to an external provider through a plugin or API.

A15.19 We do not propose to specify which approach or combination of approaches should be taken by a service when sourcing and implementing a hash database. We consider that the most important factor for the accurate and effective operation of hash matching on a service is the appropriateness of the hash database, regardless of how this is sourced.

A15.20 We therefore propose to set out some conditions that any hash database should meet in order for it to be appropriate for use by a service. These are designed to ensure that the adoption of hash matching technology is suitably effective and accurate in mitigating against the circulation of CSAM on a service. These are set out below.

Conditions for an appropriate hash database

A15.21 **Content**: Databases should at a minimum cover CSAM – i.e. content determined to be illegal, in accordance with relevant criminal law in the UK. This includes images that may not be illegal in other jurisdictions, such as drawn CSAM. This also recognises that services may use hash matching to detect other forms of content, such as contextual imagery or content that otherwise violates their terms of service.

A15.22 **Content addition**: Hashes in the database should be sourced from an organisation or person with expertise in the identification of CSAM, and the database should be regularly updated with newly-discovered content using the organisation's or persons' preferred means (for example, web crawling, human review of content, public reports).

A15.23 **Governance**: There should be governance arrangements in place to ensure that CSAM is included in the hash database correctly, and to allow for hashes of CSAM to be reviewed and removed swiftly if found to be incorrect. These governance arrangements would need to

apply as widely as necessary to ensure they operate effectively in relation to hashed CSAM (for instance, if hashes of CSAM cannot be distinguished from hashes of other material in the database, the arrangements should allow for review of all relevant hashed material). Third-party providers should preferably provide automated means for services to report such potentially incorrect hashes to maintain the accuracy of the database as far as possible.

A15.24 **Security**: The database should be secured against unauthorised access, interference or exploitation. This should include technical and non-technical measures, comprising of a mix of procedural, physical, personnel, and technical controls, to secure against adversarial attacks and exploits.

Note on internal-only databases and accessing external databases

A15.25 Our proposal recommends that hashes be sourced from an organisation or person with expertise in the identification of CSAM. We are aware that third-party providers of hash databases collate hashes from a range of sources to ensure that a wide breadth of CSAM is captured and hashed.

A15.26 Whilst we understand that some services use internal hash databases populated with CSAM identified on their service, we consider that there will be significant limitations to any hash matching system which uses only user reporting/flagging or internal detection technologies to populate their hash database. Whilst any effort to remove online CSAM is beneficial, it would be unlikely that any such database would sufficiently represent the breadth of known CSAM that is in circulation online, which would impair content detection and removal efforts. For this reason we consider that services should populate their database with hashes from third-party providers, which are more likely to contain a higher volume of content from a wider range of sources.

A15.27 However, we recognise that services may need to meet certain access criteria to get access to certain hash databases and hash matching services. Through our engagement with stakeholders, we understand access issues are more likely to arise for services with specific functionalities, or for specific types of service. To address these issues, services could:

- Work with third parties to understand why access is denied to them, and seek to resolve any issues to enable access to the database;
- Seek access to a database from another provider which may have different access criteria.

A15.28 We recognise that there may be circumstances in which services are unable to access hash databases despite addressing any potential issues. On this matter, **we welcome input and evidence from providers of third-party databases and/or hash matching services** to indicate whether the services in scope of this recommendation would be able to access their hash databases for the purpose of implementing the proposed measure.

The breadth of content scanned on the service

A15.29 We have considered the scale of application of perceptual hash matching technology for services adopting this measure.

A15.30 Given the risk of harm that is posed by this content, we consider it proportionate for services to apply perceptual hash matching technology to all publicly communicated content on their service, including:

- all regulated user-generated content present on the service at the time the measure is implemented, within a reasonable time of the technology being implemented;
- ongoing review of all new content posted to the service, which services would analyse before or as soon as is practicable after it can be encountered by another user, or other users, of the service; and
- inclusive of all still or animated images (e.g. photographs, pseudo-photographs, drawn images) and videos.

A15.31 We note that there may be additional complexity to hash matching video content, however we consider that it is still proportionate to include video hashing in the measure due to the risk of harm posed by this content.

Technical performance

A15.32 An image is detected as a match for known CSAM where the hash created from the image is considered sufficiently similar to the hash created from the known CSAM. Perceptual hash matching technology uses 'distance metrics' to measure the similarity between the hashes, and a threshold is set to determine when there is sufficient similarity to constitute a match.

A15.33 Such technical parameters are critical to the technology's accuracy and effectiveness. Put simply, a threshold set too low will be ineffective in detecting modified versions of known CSAM; while a threshold set too high will result in excessive numbers of false positives.

A15.34 Adjustments can be made to the distance metric, to how the distance between two hashes is calculated, and/or to the threshold to ensure an appropriate level of technical performance. We have therefore considered how to address this in the design of any measure.

A15.35 Different perceptual hash functions take different approaches to creating a hash from an image, and these approaches may require the use of different distance metrics. As the threshold would depend on the distance metrics used, and given that the measure discussed in this section would not specify the use of a particular hash function, it would therefore not be practicable to specify a particular threshold which should be used to determine whether an image is a match.

A15.36 We also expect services to review the performance of the system over time and that, as part of this, what is an appropriate threshold may change over time.

A15.37 In adopting this approach, the proposed recommended measure sets out that service providers should ensure the performance of the perceptual hash matching technology strikes an appropriate balance between precision and recall.[255]

A15.38 In striking that balance, service providers should take into account:

[255] Precision refers to the proportion of content detected as a match by the technology that has been correctly identified as CSAM (or other content intentionally represented in the same database); and recall refers to the proportion of content analysed by the technology which is known CSAM or a modified version of known CSAM (or other content intended to be detected) which the technology successfully detects as a match. For the avoidance of doubt, a focus solely on precision would not strike an appropriate balance as it would involve the use of an excessively low threshold for the distance metric which would fail to identify many modified versions of known CSAM. Likewise, a focus solely on recall would not strike an appropriate balance, as it would be expected to cause excessive false positives.

- the risk of harm relating to CSAM identified in the latest illegal content risk assessment of the service (including, in particular, information reasonably available to the provider about the prevalence of relevant content that is CSAM on the service);
- the proportion of content detected as a match by the technology that is a false positive; and
- the effectiveness of the systems and processes used by the service to identify false positives.

A15.39 For instance, factors which might point towards striking the balance further towards precision would be that:

- the service had assessed itself as being at medium risk of image-based CSAM, and had no information indicating that the hash matching system was generating excessive false negatives;
- a high proportion (in relative terms) of content detected by the technology (as currently configured) transpires to be false positives – demonstrated for example through the outcome of reviews of detected content by human moderators; and
- the systems and processes in place to identify false positives are relatively less effective.

A15.40 Conversely, factors which might point towards striking the balance further towards recall would be that:

- the service had assessed itself as being at high risk of image-based CSAM, and had information that known CSAM was prevalent on the service (such as from user reports or notifications from other organisations such as child protection organisations, law enforcement authorities, or other providers of internet services) but remained undetected by the hash matching system;
- a low proportion (in relative terms) of content detected by the technology (as currently configured) transpires to be false positives – demonstrated for example through the outcome of reviews of detected content by human moderators; and
- effective systems and processes are in place to identify false positives – such as review of a relatively high proportion of detected content by human moderators

A15.41 Service providers should review the performance of the system, and whether the balance between precision and recall continues to be appropriate, at least every six months. This is based on engagement with stakeholders which suggests this is an appropriate review period which aligns with current practice. We consider that regular analysis of all available performance data would help ensure the accuracy and effectiveness of the system.

A15.42 Service providers may also take other steps in response to the review, such as increasing the proportion of detected content that is reviewed by human moderators, or integrating alternative or additional hash databases.

A15.43 Service providers should ensure a written record is made of how they have struck this balance in configuring their technology, including what information has been considered, and information about reviews and the steps taken in response.[256] We consider that this would help ensure that a proper decision-making process is followed.

[256] For further information, see: Annex A6 on Record keeping and review.

Human oversight and review

A15.44 As explained above, perceptual hash matching can give rise to false positive matches for hashes from a CSAM hash database or matches for content that has been wrongly included in a hash database. The design of the elements of the measure relating to hash databases and technical performance described above help to mitigate, but do not remove, these risks.

A15.45 In light of the impact on users' rights to freedom of expression and privacy that would arise if such content is erroneously taken down, we have considered whether it would be appropriate to include provision for human review of matched content as a safeguard to mitigate those impacts. Review by human moderators is already used by many services to review at least some matches detected by perceptual hash matching technology to identify false positives.

A15.46 We do not think it would be proportionate to recommend that services review all matches detected by perceptual hash matching technology. As the technology is capable of a high degree of accuracy (dependent, for example, on the accuracy of the hash database(s) used and the configuration of the technology's technical performance), we consider that the high costs of human review of all detected content are not justified. Services may also have other measures in place to assist in identifying false positives, such as using automated content classifiers to further analyse detected content.

A15.47 We consider that services should review an appropriate proportion of content detected as a match for known CSAM. Our concern here is to reduce the amount of content that is erroneously taken down as being CSAM. If a service chooses to use hash matching to detect other kinds of content as well as CSAM, it is for it to decide what level of human review to use, if any, in relation to that content. However, if it is unable to differentiate matches for hashes of CSAM from matches for such other kinds of content, then our recommendation to use human moderators applies in relation to all detected content.

A15.48 In deciding what proportion of detected content to review, service providers should take into account the following three core principles.

A15.49 First, the amount of resource for human review should be proportionate to the degree of accuracy of the service's perceptual hash matching technology and any associated systems. This means that all else being equal we would expect services with less accurate hash matching systems to do more human review than similar services with more accurate hash matching systems. This should consider actual performance, as indicated by the results of the periodic review of the technology's performance (see paragraph A15.36 above). This review should take into account the outcomes of review by human moderators.

A15.50 Second, the resource should be targeted at content with a higher likelihood of being a false positive.

A15.51 For instance, detected content which was close to the lower bounds of any threshold used to indicate sufficient similarity to constitute a match for known CSAM would generally be more likely to be a false positive.

A15.52 Services may also use other signals to assess the likelihood that content is a false positive (for example, cryptographic hash matching could show that the content detected by perceptual hash matching is identical to content previously reviewed and assessed to be CSAM).

A15.53 Services may also choose to document hashes which are known to cause matches with benign content, and conduct human review on positive matches against those hashes.

A15.54 Third, service providers should harness the data from any user complaints relating to the removal of content through the use of perceptual hash matching for CSAM as a way of further protecting users from any incidence of false positives through the use of this technology.

A15.55 Service providers should ensure that a written record is made of their policy for review of detected content. This policy should set out the proportion of detected content which the provider plans to review, and information about how the principles set out above were taken into account in setting that policy. Again, we consider that this would help ensure that a proper decision-making process is followed.

A15.56 Service providers should also keep statistical records about content reviewed in accordance with that policy, to be used when reviewing the performance of the system. We recognise that the provider may not consistently review the proportion of detected content set in its policy – for instance, if moderation resources need to be redeployed to address a particular incident.

URL detection for CSAM

A15.57 This section discusses the design of a measure recommending the use of technology to detect and remove CSAM URLs. The outline measure based on these discussions can be found in Paragraph 14.159 of Chapter 14. A copy of the draft measure that we are proposing to include in our CSEA Code of Practice can also be found in Annex 7 to this consultation.

Type of technology

Direct matching

A15.58 Direct matching will only detect a match if a URL exactly matches that on a URL list. As this requires an exact match with previously identified URLs hosting CSAM content, the risk of surfacing links to content that is not CSAM is lower than with fuzzy matching, though this is dependent on the quality and accuracy of the underlying list.

A15.59 Our current evidence base suggests that this technology is a straightforward and accessible version of URL detection, which is likely to present fewer challenges in implementation than fuzzy matching.

Fuzzy matching

A15.60 In addition to direct matches with previously identified URLs hosting CSAM content, fuzzy matching will also surface matches that are similar (e.g. which end 'dotcom', instead of '.com'). As such, there may be benefits to using fuzzy matching rather than direct matching as it may surface more links to previously identified CSAM content. It is intended to detect matches to slightly modified URLs, and therefore may be more effective at disrupting the sharing of CSAM as a result.

A15.61 However, we have insufficient evidence on the implementation of fuzzy matching and its potential challenges. In particular, we are aware that there are likely to be potential adverse impacts on users' freedom of expression, as there is a risk of over-blocking of legitimate content. We do not currently have evidence on how accurate this technology is and therefore the materiality of this risk.

A15.62 We are therefore not proposing to recommend the use of fuzzy matching at this point, and instead we propose to focus on a potential measure recommending direct matching. Services with the capacity and capability to deploy fuzzy matching may choose to do so as this may be more effective in some circumstances, however this would not be necessary under our proposed measure at this time.

The URL list

What type of list should be used by services

A15.63 We understand that, at present, the content included and accessibility of CSAM URL lists varies between providers. For example, the IWF maintains a CSAM URL list which is made available to U2U services. There are also a small number of other third-party lists available, such as those maintained by NGOs and law enforcement bodies globally. We understand that these lists vary in terms of whether they include URLs to specific pages or domain-level URLs.

A15.64 We also understand that some lists may only be available to certain organisations, such as law enforcement bodies or internet service providers (ISPs), as opposed to U2U services. This includes, for example, Interpol's 'Worst of' list (IWOL), which is a list of domains provided to ISPs to block access at the network level.[257]

A15.65 We also note it may be possible for third party content moderation providers, with relevant expertise, to re-package URL lists as part of a wider content moderation offer.

A15.66 Further, we understand that some third party lists available to services will have access criteria, including membership, access restrictions and fees.

A15.67 A service encountering issues accessing a URL list could work with third parties to understand why access is denied to them, and seek to resolve any issues that may allow them to access the list. They could also seek access to a list from another provider which may have differing access criteria.

A15.68 We understand that some larger services may maintain their own lists of URLs. However, there are significant legal and practical difficulties arising from maintaining a list of CSAM URLs. In particular, as discussed below, maintaining a list involves putting in place arrangements to remove URLs from the list if they no longer contain CSAM. Where this is done by repeatedly accessing the URL, it could give rise to potential criminal liability (in particular for "making" an indecent image of a child, which guidance published by the Crown Prosecution Service explains has been widely interpreted by the courts). Further, the list of CSAM URLs may be more vulnerable to unauthorised access compared to, for example, hash databases where it is the hash which is stored and not the actual image.

A15.69 Given these concerns, we consider that any recommended measure should stipulate that services use lists compiled by third parties (such as certain NGOs or law enforcement authorities).

Ensuring the list is appropriately maintained

A15.70 Our provisional view is that any recommendation should provide that for a list procured from a third party to be appropriate, it should meet the following conditions:

[257] Interpol, no date. Access blocking. [accessed 24 July 2023].

- Lists should at a minimum cover URLs hosting CSAM – i.e. content determined to be illegal in accordance with relevant criminal law in the UK. This includes images that may not be illegal in other jurisdictions, such as drawn CSAM;
- The list is sourced from an organisation or persons with expertise in the identification of CSAM. The list should also be regularly updated with newly-discovered content using the organisation or person's preferred means (for example, web crawling, human review of content, public reports);
- There are governance arrangements in place to ensure that URLs at which CSAM is present are included in the list correctly, and are removed from the list if they no longer contain CSAM; and
- The list should be secured against unauthorised access, interference or exploitation, to prevent bad actors from gaining access to illegal material.

A15.71 Services should compare user generated content on their service to the most recent version of the URL list supplied by the third party.

A15.72 To ensure that URL lists are being deployed effectively and that the technology is working as intended, services should also ensure that they carry out testing and maintenance of internal systems.

The appropriate level of granularity of lists

A15.73 We understand that providers of CSAM URL lists can include URLs of specific webpages containing CSAM, or in some cases list at the domain level (e.g. a whole website).

A15.74 Listing URLs at domain level could risk users not being able to access a range of legitimate content (with implications for users' freedom of expression, and for providers of internet services listed at domain level).

A15.75 In most instances, we therefore consider that it would be appropriate for the URL to be included at the most granular level (i.e. the specific URL at which CSAM is hosted), as this addresses the harm, while reducing the risk of "over-blocking". There are however instances where listing URLs at domain level could be justified.

A15.76 We consider that it would be appropriate to list at domain level where the domain is entirely or predominantly dedicated to CSAM. This is likely to be more effective and efficient than listing each individual URL containing CSAM, given that these may alter frequently. When procuring a list from a third party, we would expect services to ensure that the third party has arrangements in place to ensure that listing at domain level only occurs in such cases.

A15.77 For the purposes of this measure, we consider that a domain is likely to be 'predominantly or entirely dedicated to CSAM' where the content present at the domain, taken overall, predominantly or entirely comprises CSAM such as indecent or prohibited images, or otherwise appears to be related to the encouragement of CSEA offences (such as advertising the distribution of CSAM). This is a qualitative rather than quantitative assessment, and would in our view include where the third-party provider assesses that the content present at the domain clearly indicates that at least some of the CSAM visible at that domain is actively uploaded, controlled or promoted by the site, or where it appears that the CSAM visible at the domain may not have been actively uploaded, controlled or promoted by the site, but there is no indication that the domain's genuine purpose is to share legal content. If CSAM is present at the top-level or homepage of the website, that would ordinarily be a strong indicator that the site can be considered to be predominantly dedicated to CSAM.

The breadth of content that is scanned

A15.78 Given the risks posed by this content, the magnitude of harm that it can cause, and the evidence indicating how widely the content can be shared across services, we consider that for any recommendation it would bring significant benefits for services to apply URL detection technology to all publicly communicated content on their service, including:

- all regulated user-generated content present on the service at the time the measure is implemented, within a reasonable time of the technology being implemented;
- ongoing review of all new content posted to the service, which services would analyse before or as soon as is practicable after it can be encountered by another user, or other users, of the service; and
- inclusive of all URLs (both as regular text and as hyperlinks).

Human review

A15.79 The appropriate level of human oversight and review will depend on the accuracy of the underlying technology and URL database. In this particular case, we note that:

- We are proposing to recommend only direct matching and removal of URLs contained on a list of URLs known and verified as CSAM (including directing to CSAM). We understand that the false positive rate for direct matching itself is low to none; and
- we would expect services to ensure that the URL lists they procure are accurately maintained and updated, and that they use the most recent version of the list made available to them.
- We therefore would not anticipate it being necessary or proportionate for services to use human review to check positive matches to the URL list before the URLs are taken down from the service. However, we do anticipate that the system itself is likely to require regular assurance reviews, for example to ensure that the list is integrated appropriately into the service's systems and to ensure that the technology is functioning as intended.

Standard keyword detection regarding articles for use in frauds

A15.80 This section discusses the design of a measure for standard keyword detection relating to illegal content that concerns articles for use in fraud. The outline measure based on these discussions can be found in Paragraph 14.234 of Chapter 14, and a copy of the draft measure that we are proposing to include in our Illegal Content Code of Practice can be found in Annex 7 to this consultation.

The type of technology

A15.81 As with URL detection for CSAM, there are two main ways that services can use standard keyword detection technology to identify potentially illegal content: direct matching and fuzzy matching.

A15.82 We recognise that the direct method of matching is more accurate as it only identifies content which contains keywords or combinations of keywords that feature on the

underlying keyword list. Any variations in the spelling of the words would result in content not being detected as a match to the keyword list. There is therefore a lower risk of false positives being identified.

A15.83 It can however be more easily avoided, including by bad actors, as it is not able to detect keywords that are similar to those on the keyword list (for example, Ofcom and Ofc0m). It is possible for services to incorporate a larger number of keywords (including common misspellings) in their keyword list in order to seek to mitigate this risk, but this can be time-consuming for services and there is a risk that it still may not be as effective as fuzzy matching.

A15.84 Fuzzy matching would include direct (and case invariant) matches but, as explained in paragraphs A15.60-62 above, can also capture words that are similar to those in the keyword list. This means that it can in principle be more effective at detecting illegal or violative content (including where deliberate evasion techniques have been used).

A15.85 This does however present a higher risk that content which is not illegal or violative is identified by the technology. We understand that human review is therefore commonly deployed to assess the accuracy of the keyword detection technology and the risk of false positives, and that services can attune the parameters of their keyword search detection technology to adjust the false positive/negative rate appropriately.

A15.86 Taking account of the potentially greater effectiveness of fuzzy matching, the very specific nature of the words often used in relation to the supply of articles for use in frauds, and the severity of the harm that can arise from the dissemination of such content, our provisional view is that fuzzy matching may be the better way to address this harm.

The steps services should take to ensure that they have access to an appropriate keyword list

A15.87 We do not consider it would be appropriate for Ofcom to prescribe the keywords that services should search for; publication of such a list by Ofcom would undermine its effectiveness, and what is appropriate for inclusion within a keyword list may differ depending on the service.

A15.88 However, the effectiveness and accuracy of keyword detection technology is intrinsically linked to the underlying keyword list. Our starting point is therefore that any Code measure regarding the use of keyword detection will need to set out in sufficient detail what is expected of in-scope services with regard to the setting up and maintenance of the keyword list. If appropriate steps are not taken at either stage, the accuracy and effectiveness of the technology would likely be compromised.

A15.89 We understand that some services set up and maintain their own list of relevant keywords and / or procure a list from a third-party provider with relevant expertise. This can be an appropriate way for services to acquire such a list if they do not have the internal capacity or consider external procurement more efficient. We recognise that both can in principle be an effective means of creating and maintaining a keyword list, and we consider that any Code recommendation regarding the use of keyword search detection technology should provide services with flexibility in this regard.

A15.90 In either case, it is important in our view that services looking to deploy keyword detection technology take appropriate steps to ensure that their keyword list:

- contains only words that could not reasonably be expected to be used on the relevant service (either on their own or in combination with other keywords on

the list) except in relation to the commission of an offence concerning articles for use in frauds[258]; and
- is sufficiently comprehensive.

A15.91 We have considered the extent to which we might be able to provide further clarity in any Code measure about what these steps might be. Our provisional view is that such steps should include:

- **The compilation of an initial keyword list**. This should be a list of words that are (either on their own or in combination with other words) unlikely to be used except in relation to the commission of an offence regarding articles for use in frauds. We are recommending that, in compiling this initial keyword list, services take account of both: (1) Information sourced from one or more persons with expertise in the identification of content relating to articles for use in frauds. This could include, for example, research carried out by the service, drawing on published academic and news articles, and keyword research tools, and/or a list of key words obtained from a third party who has a particular expertise for the purposes of detecting and identifying content likely to indicate fraudulent activity; and (2) The outcomes of reviews of content carried out by human moderators, reports by users and affected persons, and content flagged or reported by Trusted Flaggers or similar systems. This analysis may identify keywords that should be added to, or removed from, the initial keyword list.
- **The taking of appropriate measures to test the list on a reasonable sample of content communicated publicly on the service**. This can be achieved through testing of keywords and combinations of keywords on the service. We do not propose to prescribe what a reasonable sample of content is; we would expect this to depend on the volume of content present on the service.
- **The removal of words from the initial keyword list** which the provider cannot reasonably expect, in light of b), to be used on the service (either on their own or in combination with other keywords on the list, as appropriate) only in relation to the commission of an offence concerning articles for use in frauds. For example, because they are used legitimately in other contexts beyond the making or supply of articles for use in frauds.
- **The taking of appropriate measures to secure the keyword list from unauthorised access, interference or exploitation** (whether by persons who work for the provider of the service or any other person).

A15.92 We think it is important that service providers make and keep a written record of the steps they took to compile their keyword list (including the information they considered and what measures they took to test the list on a reasonable sample of content), as well as what keywords they included in that list. This should help ensure that a proper decision-making process is followed.

A15.93 We are also proposing that services should review their keyword list at least every six months and that, when doing so, they should take account of:

- information sourced from one or more persons with expertise in the identification of content that amounts to an offence concerning articles for use in frauds and which has not previously been taken into account by the service. This could

[258] The fact that those words might also be used in academic or news articles discussing the supply of articles for use in frauds should not be sufficient by itself to justify the exclusion of those words from the keyword list.

include, for example, updates provided by a third party with particular expertise, and/or updated research drawing on published academic and news articles and keyword research tools;

- evidence reasonably available to the provider on the accuracy and effectiveness of the fuzzy keyword detection technology in detecting content which amounts to an offence concerning articles for use in frauds. We would expect this to include evidence from the outcomes of reviews of content carried out by human moderators, reports by users and affected persons, and content flagged or reported by Trusted Flaggers or similar systems (for example, where user reports and Trusted Flaggers report content that has not been detected by the keyword detection technology).

A15.94 Our provisional view is that reviews of the keyword list (and performance of the keyword detection technology, discussed later) should be conducted at least every six months to help ensure the accuracy and effectiveness of the technology. We would however expect the review period to be more frequent where there is evidence to suggest that this would be proportionate. Such evidence includes, for example, that reviews of detected content by human moderators are identifying a relatively large volume of false positives, or there is evidence that large volumes of illegal content concerning articles for use in frauds are not being detected by the keyword detection technology. Our provisional view reflects our understanding of industry practice, although we recognise that our evidence base is more limited in this regard and we are therefore keen for feedback and evidence from stakeholders on what is likely to be an appropriate period between reviews.

A15.95 We think it is important that service providers make and keep a written record about their review of the keyword list. This should include the date of each review, what information has been considered as part of the review, what measures were taken by the service to test any new keywords on the service, and the steps taken in response (such as which words were removed and inserted, if any, following that review and why). This should help ensure that a proper decision-making process is followed.

The breadth of content that is scanned

A15.96 We have considered the scale of application of standard keyword search detection technology for the measure under consideration.

A15.97 Given the risk of harm that is posed by this content, we are minded to suggest that standard keyword search detection technology be applied to all text-based user-generated content communicated publicly on a service. In particular, to:

- user-generated text-based content present on the service at the time the measure is implemented, within a reasonable time; and
- all new user-generated text-based content generated on, uploaded to or shared on the service after the technology has been implemented, before or as soon as is practicable after it can be encountered by another user, or other users, of the service.

A15.98 This measure would not apply to content communicated privately.

The technical performance of the technology

A15.99 The technical parameters for standard keyword detection technology can be adjusted over time, and different approaches taken by services. These parameters (alongside the keyword list itself) will impact the accuracy and effectiveness of the technology. For example:

A15.100 For all types of keyword matching:

a) the length of the string of text over which the search is conducted may vary; and
b) the number of keywords that need to be identified within a string of text (and their proximity to each other) in order for that content to be flagged to the service may vary. We refer to this as the "text density". Whilst in some cases, it may be sufficient for one or two keywords to be used within a string of text to identify potentially illegal or violative content, some services may look for a greater concentration of keyword terms per unit of text.

A15.101 For fuzzy matching specifically:

a) the length of the string of text over which the search is conducted may vary;
b) the service will need to decide what similarities should be within scope of the search. For example, some fuzzy matching algorithms only consider the difference in the manner in which words are spelt,[259] whilst others help to detect words that are spelled differently and sound similar[260];
c) the service will also need to set a 'similarity threshold'. This indicates how similar two strings of text need to be in order for content to be identified as potentially illegal or violative. The minimum value of 0.00 causes all values to match each other. The maximum value of 1.00 only allows exact matches.

A15.102 We have considered whether it would be appropriate to prescribe any of the above in setting out a keyword detection recommendation, and what recommendations it may be appropriate to make regarding the quality assurance and review of the technology's performance over time.

A15.103 For the reasons set out below, our provisional view is that it would not be appropriate to prescribe any of the above in setting out a keyword detection recommendation.

A15.104 As explained above, our research suggests that the dense combining of keywords is often used to indicate articles for use in fraud online. We recognise however that there may be some instances where a service provider is satisfied that a single keyword, when used on its service, cannot reasonably be expected to be used on its service for any purpose other than in relation to articles for use in fraud. We also recognise that a recommendation that services only search for combinations of keywords might undermine the effectiveness of the technology, indicating to bad actors that a single but effective keyword can evade keyword search detection technology. We do not therefore consider it would be appropriate to prescribe the density of keywords that the technology should search for.

---

[259] For example, the Levenshtein distance gives a measure of the number of single insertions, deletions or substitutions required to change one string into another. Source: National Institute of Standards and Technology, 2019. Levenshtein distance (nist.gov) [accessed 1 August 2023].
[260] For example, the Soundex algorithm and Metaphone algorithm. Sources: National Institute of Standards and Technology, 2021. soundex (nist.gov) [Accessed 1 August 2023] and National Institute of Standards and Technology, 2022. metaphone (nist.gov) [Accessed 1 August 2023].

A15.105 Similarly, we do not consider it would be appropriate to prescribe the string length, fuzzy matching algorithm or similarity threshold to be used by services. We recognise that

a) what might be an appropriate string length may differ depending on, for example, the service, the underlying keyword list, and the similarity threshold;
b) different fuzzy matching algorithms take different approaches to identifying whether a string of text contains keywords (or text that is sufficiently similar to keywords) and that, in principle, these can all be effective (and that some services may even choose to use multiple fuzzy matching algorithms as part of a layered keyword search detection proposition, to improve its effectiveness); and
c) what is an appropriate similarity threshold will vary depending on, for example, the fuzzy matching algorithm, the string length, the effectiveness of any systems and processes that the service uses to identify false positives and the underlying keyword list. For example, a lower similarity threshold may be appropriate where a keyword list does not contain common misspellings or alternative spellings for keywords.

A15.106 We also expect services to review the configuration of their technology over time and that, as part of this, what is an appropriate fuzzy matching algorithm, similarity threshold or string length may change over time.

A15.107 We therefore consider, similar to our CSAM hash matching measure, that any measure regarding the use of keyword search detection technology to identify content that is likely to amount to a priority offence concerning articles for use in frauds should instead provide that service providers ensure the performance of the keyword search detection technology strikes an appropriate balance between precision and recall.

A15.108 A focus solely on precision would not strike an appropriate balance as it would involve the use of an excessively high threshold for similarity. Conversely, where a keyword has been added to a service's keyword list because, in combination with other keywords on the list, that service provider considers it would not be used for any purpose other than in relation to articles for use in frauds, a search for that one keyword by itself would not be consistent with the service striking an appropriate balance between precision and recall.

A15.109 In striking that balance, service providers should take into account

a) the risk of harm relating to fraud identified in the latest illegal content risk assessment of the service (including in particular information reasonably available to the provider about the prevalence of relevant content which amounts to an offence concerning articles for use in frauds on the service);
b) the proportion of content detected by the keyword search detection technology that is a false positive, which we discuss further below; and
c) the effectiveness of any systems and processes used by the service to identify false positives before content is taken down.

A15.110 For instance, factors which might point towards striking the balance further towards precision would be that:

a) the service had no information indicating that content which amounts to an offence concerning articles for use in frauds was prevalent on the service;
b) a high proportion (in relative terms) of content detected by the technology (as currently configured) transpires to be false positives – demonstrated for example through the outcome of reviews of detected content by human moderators; and

c) there are no systems and processes in place to identify false positives before content is taken down, or such systems are relatively less effective.

A15.111 Conversely, factors which might point towards striking the balance further toward recall would be that:

a) the service had information that relevant content which amounts to an offence concerning articles for use in frauds was prevalent on the service (such as from user reports or notifications from other organisations, such as fraud awareness organisations, law enforcement authorities, or other providers of internet services);
b) a low proportion (in relative terms) of content detected by the technology (as currently configured) transpires to be false positives; and
c) any systems and processes in place to identify false positives before content is taken down are relatively more effective (for example, review of a relatively high proportion of detected content by human moderators).

A15.112 We are proposing that service providers should review the performance of the technology, and whether the balance between precision and recall continues to be appropriate, at least every six months (and at the same time as they review their keyword list). Regular review should help ensure the accuracy and effectiveness of the technology. We recognise that service providers may take other steps in response to the review of the technology's performance, such as increasing the effectiveness of their systems and processes to identify false positives before content is taken down, or updating their keyword list.

A15.113 Our proposal that review of the keyword list and technology's performance be conducted in parallel reflects that both are fundamental to the accuracy and effectiveness of the technology (and our proposal that, in reviewing the keyword list, services should consider evidence on the accuracy and effectiveness of the keyword detection technology more generally).

A15.114 Service providers should ensure a written record is made of how they have struck this balance in configuring their technology, including what information has been considered, and information about reviews and the steps taken in response. We consider that this would help ensure that a proper decision-making process is followed.

Human oversight

A15.115 We understand that services which already deploy keyword search detection commonly use some form of human review as a form of quality assurance to assess the accuracy of the keyword detection technology and underlying keyword list. We have therefore considered to what extent this should form part of any recommendation on the use of keyword detection technology. Some services might also use human review before any decision is made about whether to take content identified by the keyword detection technology down (and we discuss this separately at paragraph A15.120 below).

A15.116 Our provisional view is that human review of a reasonable sample of content identified by the technology can be important to ensure that the technology and underlying keyword list are accurate and effective, so far as possible. It provides the service with information about the proportion of content detected by the technology that is a false positive, which can inform its reviews of the keyword list and technical performance of the technology. Specifically, it might identify words or combinations of keywords that should be

removed from the keyword list and/or indicate that the provider should consider reconfiguring the technology's parameters.

A15.117 We are therefore proposing that a reasonable sample of detected content should be reviewed by human moderators in each review period. Whilst we are not minded to prescribe what we consider a reasonable sample to be, we are proposing that services should take account of the following principles when doing so

a) What is a reasonable sample size within a period should be decided having regard to:

- The volume of content detected by the technology since the last review of the technology and keyword list. The more content that is identified by the technology, the larger we would expect the sample to be.
- The volume of false positives identified by human moderators in the preceding review period. For instance, a greater volume of detected content should be reviewed by human moderators if the technology's technical performance was found in the last review period to give rise to a relatively large volume of false positives; and
- Data from the complaint procedure enabling users to complain if they believe their content has been wrongly identified as illegal content since the last review of the technology and keyword list.

A15.118 Whilst the sample should be targeted primarily at content with a higher likelihood of being a false positive, it should also target some content identified as having a lower likelihood of being a false positive.

A15.119 We are proposing that service providers should ensure that a written record is made of the volume of content which has been reviewed by human moderators in each review period, the proportion of such content that is a false positive, and information about how the principles set out above were taken into account in setting that policy.

What steps should services take in relation to detected content?

A15.120 We have set out above a number of recommendations that are designed to secure the accuracy of the keyword detection technology, so far as possible, in correctly identifying content amounting to the supply of articles for use in frauds. We would expect any content detected by technology deployed in accordance with our measure to be highly likely to be illegal content.

A15.121 However, we recognise that the keyword detection measure we are considering in relation to the supply of articles for use in frauds will enable services to identify content that is likely to be illegal, but about which no prior illegal content judgement or determination has been made. It may identify legitimate content (such as news articles or academic articles) which discuss the commission of offences regarding articles for use in fraud. It therefore differs from our proposed measures regarding CSAM hashing and the detection of CSEA links which focus on the detection of positive matches with content (or URLs that provide access to content) that has already been determined to be illegal.

A15.122 In light of this, we do not consider it appropriate to recommend that services swiftly take down all content detected as a positive match by their keyword detection technology, instead we recommend (as discussed below) that the decision on whether or not the

content should be taken down should be taken in accordance with their content moderation systems and processes.[261]

A15.123 Consistent with Chapter 12, we are not persuaded that it would be appropriate to specify in detail how services should configure their content moderation systems and processes to take account of content detected by the keyword detection technology (for example, that there be human moderation of all such content), or the outcomes that those systems and processes should achieve (for example, through detailed KPIs).

A15.124 We are proposing in that Chapter that all U2U service providers must have in place content moderation systems or processes designed to take down illegal content swiftly.

A15.125 We are also proposing in that chapter to recommend the following measures to all multi-risk U2U services and all large U2U services:

a) Services should set internal content policies having regard to the findings of their risk assessment and any evidence of emerging harms on their service;
b) Services should set performance targets for their content moderation functions and measure whether they are achieving these;
c) Services should have and apply policies on prioritising content for review;
d) Services should resource their content moderation functions so as to give effect to their internal content policies and performance targets;
e) Staff working in content moderation must receive training and materials to enable them to identify and take down illegal content.

A15.126 Our provisional view is that any recommendation that services use keyword detection technology to identify content likely to amount to a priority offence regarding articles for use in frauds should specify that detected content be considered in accordance with that service's internal content policies. This should provide flexibility to services to determine how to moderate detected content, and (subject to the below) how to prioritise it as against other potentially illegal content (for example, content suspected to be illegal based on user reports or complaints).

A15.127 As part of our wider proposals on content moderation, we explain that - when prioritising what content to review, large, multi-risk or single risk services should at a minimum have regard to the following factors: virality of content, potential severity of content, the likelihood that content is illegal (including whether it has been flagged by a trusted flagger). We would expect any content detected by the use of keyword detection technology deployed in accordance with our measure to be highly likely to be illegal content, and for services to therefore have regard to this when prioritising what content to review.

[261] For further information, see: Chapter 12 on Content Moderation.

A16. Glossary

This glossary defines the terms we have used throughout the consultation.

Term[262]	Definition
2020 Video-Sharing Platform Regulation Call for Evidence	'Video-sharing platform regulation Call for Evidence', published by Ofcom on 16 July 2020 , available at Call for evidence: Video-sharing platform regulation (ofcom.org.uk).
2022 Illegal Harms Call for Evidence	'*First phase of online safety regulation Call for Evidence*', published by Ofcom on 6 July 2022, available at Call for evidence: First phase of online safety regulation (ofcom.org.uk).
Act	The Online Safety Act 2023.
Adult services	A user-to-user service type describing services that are primarily used for the dissemination of user-generated adult content.
Aggregators	Services which gather clips from external services by an automated tool. They embed or link to content hosted on other services, rather than publishing their own content or hosting content uploaded by users.
Ancillary Service	A service which facilitates the provision of a regulated service (or part of it), whether directly or indirectly, or displays or promotes content relating to the regulated service (or to part of it).
Appeals (Search)	A complaint by an interested person if the provider of a search service takes or uses measures in order to comply with the illegal content safety duties, that result in content relating to that interested person no longer appearing in search results or being given a lower priority in search results.
Appeals (U2U)	A complaint by a user about any of the following actions, if the action concerned has been taken by the provider on the basis that content generated, uploaded or shared by the user is illegal content: a. the content being taken down; b. the user being given a warning; c. the user being suspended, banned, or in any other way restricted from using the service.
Audio Sharing Services	Audio sharing services typically enable users to share, store, and listen to audio files such as music, podcasts, and voice recordings.

[262] The offences listed in this glossary are based on the priority offences specified in Schedules 5, 6 and 7 and certain relevant non-priority offences. For each of them,, we have provided a short explanation of the offence(s) in layman's terms. This is because we consider it important for our glossary to aid accessibility of the terms used in this document. They will not fully reflect the definition included in relevant legislation. As such, for those stakeholders who are interested in the detailed legal definition of the offence(s), please refer to the text or footnotes, which signal where the offence(s) is set out in legislation. In each case, the offence includes any associated inchoate offences.

Block	A U2U functionality where: a) blocked users cannot send direct messages to the blocking user and vice versa; b) the blocking user will not encounter any content posted by blocked users on the service and vice versa; c) the blocking user and blocked user, if they were connected, will no longer be connected.
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.
CA 2003	The Communications Act 2003.
Clear Web	Publicly accessible websites that are indexed by search engines.
Codes of practice (Codes)	The set of measures recommended for compliance with the illegal content safety duties and reporting and complaints duties that Ofcom is required to prepare under section 41 of the Act. The draft U2U and search codes on which we are consulting can be found under annex 7 and 8 of this document (respectively).
Controlling or Coercive Behaviour (CCB)	Behaviour associated where the victim and 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.[263]
Combined Service	A regulated U2U service that includes a public search engine.
CSAM (child sexual abuse material)	A category of CSEA content, including in particular indecent or prohibited images of children (including still and animated images, and videos, and including photographs, pseudo-photographs and non-photographic images such as drawings). CSAM also includes other material that includes advice about grooming or abusing a child sexually or which is an obscene article encouraging the commission of other child sexual exploitation and abuse offences. Furthermore, it includes content which links or otherwise directs users to such material, or which advertises the distribution or showing of CSAM.
CSAM URL	A URL at which CSAM is present, or which includes a domain which is entirely or predominantly dedicated to CSAM, (and for this purpose a domain is "entirely or predominantly dedicated" to CSAM if the content present at the domain, taken overall, entirely or predominantly comprises CSAM (such as indecent images of children) or content related to CSEA content).
CSEA (child sexual exploitation and abuse)	Refers to offences specified in Schedule 6 of the Act, including offences related to CSAM and grooming. CSEA includes but is not

[263] An offence under section 76 of the Serious Crime Act 2015.

	limited to causing or enticing a child or young person to take part in sexual activities, sexual communication with a child and the possession or distribution of indecent images.
Cyberflashing	The sending of a photograph or film of genitals, intending the recipient will be caused alarm, distress or humiliation, or sending a photograph or film of genitals to obtain sexual gratification and being reckless as to whether the recipient will be caused alarm, distress or humiliation.[264]
Cybermobbing	Refers to more than one person directing abusive comments towards an individual online.
Cyberstalking	Commonly used to refer to harassment and stalking taking place through electronic means, such as the internet.
Dating services	Dating services enable users to find and communicate with romantic or sexual partners.
Dedicated Reporting Channel (DRC)	A means for a Trusted Flagger (defined below) to report problems, for example an inbox, a web portal or another relevant mechanism for reporting.
Deindexing (or delisting)	Action taken by a search service, where relevant, which involves removing a URL from a search index such that it can no longer be presented to users in search results.
Digital Services Act	Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC.
Discussion forums and chat rooms	Discussion forums and chat rooms generally allow users to send or post messages that can be read by the public or an open group of people. Spoken or written communication in chat rooms typically takes place in real time, whereas posting messages in discussion forums does not.
Direct messaging	User-to-user service functionality that allows a user to send and receive a message to one recipient at a time and which can only be immediately viewed or read by that specific recipient.
Downranking	Action taken by a search service which involves altering the ranking algorithm such that a particular piece of search content appears lower in the search results and is therefore less discoverable to users.
Downstream general search services	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.

[264] An offence under section 66A of the Sexual Offences Act 2003.

Drugs and psychoactive substances offences	The supply or offer to supply of controlled drugs and/or psychoactive substances, and related offences.[265]
EA 2010	The Equality Act 2010.
ECHR	The European Convention on Human Rights (incorporated into domestic law by the Human Rights Act 1998).
Encouraging or assisting suicide or serious self-harm	When an individual intentionally encourages or assists another person to seriously self-harm or either end their life or attempt to end their life.[266]
Enforcement guidance	The guidance on how Ofcom proposes to exercise its enforcement functions that Ofcom is required to produce under section 151 of the Act. The draft guidance on which we are consulting can be found under annex 11 of this document.
Epilepsy trolling	Involves sending flashing images electronically to trigger seizures, or cause alarm or distress, among people with epilepsy, such as those with photosensitive epilepsy.
External content policies	Publicly available documents aimed at users of the service which provide an overview of a service's rules about what content is allowed and what is not. These are often in the form of terms of service and/or community guidelines.
Extreme pornography	An umbrella term to cover several categories of images which are illegal to possess, broadly covering images which are produced principally for sexual arousal, and which depict extreme or obscene behaviours.[267]
File-storage and file-sharing service	A service whose primary functionalities involve enabling users to (i) store digital content, including images and videos, on the cloud or dedicated server(s); and (ii) share access to that content through the provision of links (such as unique URLs or hyperlinks) that lead directly to the content for the purpose of enabling other users to encounter or interact with the content.
Firearms and other weapons offences	These offences relate to a wide variety of offences such as, but not limited to the purchase and sale of prohibited weapons, supplying firearms and imitation firearms to minors, purchase of firearms or ammunition without a certificate etc.[268]

[265] An offence under: section 4(3), 9A, or 19 of the Misuse of Drugs Act 1971; section 5 of the Psychoactive Substances Act 2016.

[266] An offence under: section 2 of the Suicide Act 1961; section 13 of the Criminal Justice Act (Northern Ireland) 1966 (c.20 (N.I.)); section 184 of the Act (a relevant non-priority offence).

[267] An offence under section 63 of the Criminal Justice and Immigration Act 2008.

[268] An offence under: sections 1(1) or (2) of the Restriction of Offensive Weapons Act 1959; sections 1(1), 2(1), 3(1), 3(2), 5(1), 5(1A), 5(2A), 21(5), 22(1), 24, or 24A of the Firearms Act 1968; sections 1, or 2 of the Crossbows Act 1987; sections 141(1), 141(4), or 141A of the Criminal Justice Act 1988;

Foreign Interference Offence (FIO)	Malign activity carried out for, or on behalf of, or intended to benefit, a foreign power.[269]
Fraud and financial services offences	A number of offences relating to fraud and financial services, such as but not limited to fraud by abuse of position, participating in fraudulent business, or the contravention of the prohibition on carrying on regulated activity unless authorised or exempt.[270]
Fundraising service	A fundraising service typically enables users to create fundraising campaigns and collect donations from users.
Gaming service	A gaming service allows for user-to-user interaction in partially or fully simulated virtual environments.
General search service	A service that enables users to search the internet by inputting search requests. It derives search results from an underlying search index (developed by either the provider of the service or a third party). Search results are presented using algorithms that rank based on relevance to a search request.
Geo-tagging	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.
Grooming	An offence under paragraphs 5, 6, 11 or 12 of Schedule 6 to the Act, including but not limited to the act of an abuser communicating with a child.
Harassment, stalking threats, and abuse	A range of behaviours that can cause, for example, alarm and distress to other individuals, or put them in fear of violence.[271]

articles 53 or 54 of the Criminal Justice (Northern Ireland) Order 1996 (S.I. 1996/3160 (N.I. 24)); sections 1, or 2 of the Knives Act 1997; articles 24, 37(1), 45(1) or (2), 63(8) or 66A of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3)); sections 36(1)(c) or (d) of the Violent Crime Reduction Act 2006; sections 2 or 24 of the Air Weapons and Licensing (Scotland) Act 2015 (asp 10).

[269] An offence under section 13 of the National Security Act 2023.

[270] An offence under: sections 2, 4, 7, or 9 of the Fraud Act 2006; section 49(3) of the Criminal Justice and Licensing (Scotland) Act 2010; sections 23, 24, or 25 of the Financial Services and Markets Act 2000;sections 89 or 90 of the Financial Services Act 2012.

[271] An offence under: section 16 of the Offences against the Person Act 1861; sections 4, 4A, or 5 of the Public Order Act 1986; sections 2, 2A, 4, or 4A of the Protection from Harassment Act 1997; article 4s or 6 of the Protection from Harassment (Northern Ireland) Order 1997 (S.I. 1997/1180 (N.I. 9)); sections 38 or 39 of the Criminal Justice and Licensing (Scotland) Act 2010 (asp 13).

Hate offences	Public order offences relating to stirring up hatred on the grounds of certain protected characteristics.[272]
Identity Verification (IDV)	The process of a service confirming that a user is the person they claim to be or possess an attribute they claim to have. Levels of assurance vary from service to service and the method they use to verify identity.
Illegal content	Content which amounts to a relevant offence.
Illegal content judgement guidance (ICJG)	The guidance about making illegal content judgements that Ofcom is required to produce under section 193 of the Act. The draft ICJG on which we are consulting can be found under chapter 26 of this document.
Illegal content proxy	For U2U, content that has been assessed and identified as being in breach of the service's terms of service, where the provider is satisfied that the terms in question prohibit the types of content that include illegal content (including but not limited to priority illegal content). For search, search content that is content of a kind that: a) is not allowed by the service's internal content policies, where the provider is satisfied that illegal content is included within that kind of content (including but not limited to priority illegal content); or b) contravenes the service's publicly available statement, where the provider is satisfied that illegal content is included within that kind of content (including but not limited to priority illegal content).
Illegal content safety duties	The duties in section 10 of the Act (U2U services) and section 27 of the Act (search services).
Illegal harm	Harms arising from illegal content and the commission and facilitation of priority offences.
Image-based CSAM	CSAM in the form of photographs, videos, or visual images.
Inchoate offences	Includes encouraging, assisting, conspiring to commit, aiding, abetting, counselling, procuring, attempting, or (in Scotland), inciting or being involved art and part in the commission of an offence.
Information sharing services	Information sharing services are primarily focused on providing user-generated information to other users.
Internal content policies	More detailed versions of external content policies which set out rules, standards or guidelines, including around what content is allowed and what is not, as well as providing a framework for how policies should be operationalised and enforced.

[272] An offence under: sections 18, 19, 21, 29B, 29C, or 29E of the Public Order Act 1986; sections 31 or 32 of the Crime and Disorder Act 1998; section 50A of the Criminal Law (Consolidation) (Scotland) Act 1995.

Internet Referral Units	Government-established entities responsible for flagging content to internet platforms that violates the platform's Terms of Service.
Intimate image abuse	Sharing or threatening to disclose intimate images without consent with intent to cause humiliation, alarm or distress, or for obtaining sexual gratification.273
Large service	A service with more than 7 million monthly UK users.
Low-risk service	A service that has assessed itself as being at low risk for all kinds of harm in its risk assessment.
Marketplace and listing services	Online marketplaces and listing services allow users to buy and sell goods or services.
Messaging service	Messaging services enable users to send and receive messages that can only be viewed or read by a specific recipient or group of people.
Micro-businesses	Businesses that employ 1-9 full-time employees.
Monetised scheme	A scheme by which a service labels the account of a user who has made payment to the provider of the service or some other person. Such schemes may be open to all users and payment may be regular or one-off. Users participating in the scheme may benefit from access to additional features on the service. The label to indicate that a user is participating in a paid scheme may appear on that user's profile page and/or any content they publish. Services may or may not refer to such schemes as "verification" schemes.
Multi-risk service	A service that assesses itself as being at medium or high risk in relation to at least two different kinds of illegal harm in their latest illegal harms risk assessment.
Muting	A user tool that enables a user to 'mute' another user. The muting user will not encounter any content posted by muted users on the service (unless the muting user visits the user profile of the muted user directly). The muted user is not aware that they have been muted and continues to encounter content posted by the muting user.
Network expansion functionalities	A functionality operated by means of a network recommender system which recommends other users to connect with, based on what the service knows about its users. This can include specific users who have similar interests, who are close

[273] An offence under: section 33 of the Criminal Justice and Courts Act 2015 (if the offence set out in section 188 of the Act is brought into force and Schedule 7 to the Act is amended accordingly before we issue our final document, this would refer instead to section 66B of the Sexual Offences Act 2003); section 2 of the Abusive Behaviour and Sexual Harm (Scotland) Act 2016.

	geographically, who attend the same school or workplace, or with whom a user has a mutual connection.
News publisher content	Content generated directly on a service by a recognised news publisher, or uploaded or shared on a service by a user of that service in its entirety or as a link to the original content.
Notable user scheme (or notable user verification scheme)	A scheme by which a service labels the user profile of a user to indicate to other users that they are notable. "Notable users" include but are not limited to politicians, celebrities, influencers, financial advisors, company executives, journalists, government departments and institutions, non-governmental organisations, financial institutions, media outlets, and companies. The label to indicate that a user is notable (for example a "tick" symbol) may appear on that user's user profile and/or any content they publish. Services may or may not refer to such schemes as "verification" schemes.
On-platform testing	The process of testing two or more variants of a recommender system before proceeding with a design change. During testing, services collect data that can then be used to produce metrics relating to certain identified factors, such as commercial or user safety. On-platform tests can involve different methods and are set up and executed in a testing environment.
Predictive search functionality	An algorithmic feature that is embedded in the search field of some search services. When a user begins to input a search request, the algorithm predicts the search and suggests possible related search terms. Predictions are based on many factors including past and other user queries, location and trends.
Priority illegal content	Content which amounts to a priority offence.
Priority offences	The offences set out in Schedules 5 (Terrorism offences), 6 (CSEA offences) and 7 (Priority offences) to the Act.
Proactive technology	Consisting of three types of technology: content identification technology, user profiling technology, and behaviour identification technology (subject to certain exceptions) as defined in section 231 of the Act.
Proceeds of crime	Offences relating to money or assets gained by criminals during their criminal activity, including money laundering.[274]
Product	An all-encompassing term that includes any functionality, feature, tool, or policy that a service provides to enable users to interact with or use the service.
Provider content	Any content that is published on a service by the service provider or someone acting on their behalf.

[274] An offence under: sections 327 328, or 329 of the Proceeds of Crime Act 2002,.

Provider pornographic service	An internet service on which pornographic content (defined in the Act as 'regulated provider pornographic content') is published or displayed by the provider of the service.
Publicly Available Statement	A statement that search services are required to make available to members of the public in the UK, often detailing various information on how the service operates.
Recommender system	An algorithmic system which, by means of a machine learning model, determines the relative ranking of an identified pool of user-generated content on content feeds such as newsfeeds and reels. Content is recommended based on factors that it is programmed to account for, such as popularity of content, characteristics of a user, or predicted engagement.
Record keeping and review guidance	The guidance that Ofcom is required to produce under section 52(3) of the Act to help services to comply with their record keeping and review duties under sections 23 (U2U) and 32 (search) of the Act. The draft guidance on which we are consulting can be found under annex 6 of this document.
Register of risks	The assessment of the risks of harm from illegal content on U2U and search services that Ofcom is required to prepare under section 98 of the Act. It can be found under chapter 6 of this document.
Relevant non-priority illegal content	Content which amounts to a non-priority offence.
Relevant non-priority offence	Offences under UK law which are *not* priority offences under Schedules 5, 6 or 7 to the Act, where: a. The victim or intended victim of the offence is an individual (or individuals); b. The offence is created by the Online Safety Act, another Act, an Order in Council or other relevant instrument. The effect of this is that offences created by the UK courts are not relevant non-priority offences, and offences created in the devolved Parliaments or Assemblies are only relevant non-priority offences if certain procedures are followed in their making; 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.
Relevant offences	All priority offences and relevant non-priority offences.
Reporting and complaints duties	The duties in sections 20 and 21 of the Act (U2U services) and sections 31 and 32 of the Act (search services).
Review service	A service which enables users to create and view critical appraisals of people, businesses, products, or services.

Risk assessment	The assessment required to be carried out by a service under section 9 of the Act (U2U services) or section 26 of the Act (search services).
Risk Assessment Duties	The duties under section 9 of the Act (U2U services) and section 26 of the Act (search services).
Risk assessment guidance	The guidance to assist services in complying with the risk assessment duties that Ofcom is required to produce under section 99 of the Act. The draft guidance on which we are consulting can be found under annex 5 of this document.
Risk profiles	Prepared under section 98 of the Act and as set out in Appendix A of the Illegal Content Risk Assessment Guidance.
Safe search	A feature of several general search services which filters or obscures certain kinds of search content, such as pornographic/sexual or violent content. Safe search features can have levels or can be opted in or out of. In some cases, a safe search feature is enabled by default, for example for children.
Search content	Content that may be encountered in or via search results of a search service. It does not include paid-for advertisements, news publisher content, or content that reproduces, links to, or is a recording of, news publisher content.
Search engine	Includes a service or functionality which enables a person to search some websites or databases but does not include a service which enables a person to search just one website database.
Search index	A collection of URLs that are obtained by deploying crawlers to find content across the web, which is subsequently stored and organised.
Search results	Content presented to a user of a search service by operation of the search engine in response to a search request made by the user.
Search service	An internet service that is, or includes, a search engine.
Service	A regulated user-to-user or search service, i.e. only the U2U or search part of the service. We also use it as a shorthand way of referring to the provider of the service concerned.
Service restriction order	An order that requires 'ancillary providers', such as search engines and payment services which facilitate the provision of the service, to take steps aimed at disrupting the non-compliant service's business in the UK. These orders can also be made on a temporary (interim) basis.

Sexual exploitation of adults offences	Causing or inciting prostitution for gain, or controlling a prostitute for gain.[275]
Smaller service	A service which is not a large service.
Small business	A business that employs 10-49 full-time employees.
Social media service	Social media services connect users and enable them to build communities around common interests or connections.
Specific-risk service	A service which has assessed itself as being at medium or high risk for a specific kind of harm for which we propose a particular measure.
Super-complaint	A complaint made under section 170 of the Act.
Takedown duty	The duty under section 10(3)(b) of the Act for a U2U service to use proportionate systems and processes designed to swiftly take down any (priority or non-priority) illegal content when it becomes aware of it.
Terms of Service	All documents comprising the contract for use of the service (or of part of it) by United Kingdom users.
Terrorism	An offence specified in Schedule 5 to the Act, including but not limited to offences relating to proscribed organisations, encouraging terrorism, training and financing terrorism.
Trusted Flagger	Individuals, NGOs, government agencies, and other entities that have demonstrated accuracy and reliability in flagging content that violates a platform's Terms of Service. As a result, they often receive special flagging tools such as the ability to bulk flag content.
U2U	Shorthand for 'user-to-user' service, which means 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.
Unlawful immigration and human trafficking offences	Offences relating to illegal entry, assisting unlawful immigration, or arranging or facilitating the travel of another person, or taking a relevant action, with a view to them being exploited.[276]
URL (Uniform Resource Locator)	A "uniform resource locator", which is a reference that specifies the location of a resource accessible by means of the internet.

[275] An offence under: section 52 or 53 of the Sexual Offences Act 2003; articles 62 or 63 of the Sexual Offences (Northern Ireland) Order 2008 (S.I. 2008/1769 (N/I. 2)).

[276] An offence under: sections 24(A1), (B1), (C1) or (D1) or 25 of the Immigration Act 1971; 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.)).

User-generated content	Content (a) that is: (i) generated directly on the service by a user of the service, or (ii) uploaded to or shared on the service by a user of the service, and (b) which may be encountered by another user, or other users, of the service by means of the service.
Vertical search service	A search service that enables 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 service	A service that allows users to connect and upload and share videos with the public.