Protecting people from illegal harms online

Annex 7: Illegal Content Codes of Practice for user-to-user services

DRAFT FOR CONSULTATION

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

Contents

Annex

A1. Introduction

Regulatory and legal framework

A1.1 Under the Online Safety Act 2023 (the '**Act**'), Ofcom is required to prepare and issue Codes of Practice ('**Codes**') for providers of Part 3 services, describing measures recommended for compliance with specified duties. This document contains Codes applying to the providers of regulated user-to-user services (including to the providers of combined services to the extent that the duties applicable to user-to-user services apply to those services).

A1.2 These Codes are for the purpose of compliance with the following duties:

a) the illegal content safety duties (section 10);
b) so far as it relates to illegal content, the duty about content reporting (section 20); and
c) so far as it relates to the complaints set out in section 21(4), the duties about complaints procedures (section 21).

A1.3 Ofcom must carry out its functions compatibly with the Human Rights Act 1998, including the rights to freedom of expression and privacy. In particular, any limitation on the right to freedom of expression must be prescribed by law, pursue a legitimate aim and be necessary in a democratic society. Any limitation on the right to privacy must be in accordance with the law, pursue a legitimate aim and be necessary in a democratic society. In order to be 'necessary', the restriction must correspond to a pressing social need, and it must be proportionate to the legitimate aim pursued. The legitimate aims pursued include national security, the prevention of crime, the protection of health and morals, and the protection of the rights of others. Ofcom has had careful regard to these rights in the development of our Codes, both in terms of proportionality and appropriate safeguards around users' freedom of expression and privacy.

A1.4 Ofcom is preparing Codes in three phases, as relevant to the full list of duties set out in section 41(10) of the Act. Recommendations for the purpose of compliance with the children's online safety duties (section 12) will be found in our Protection of Children Codes, while user empowerment (section 15), content of democratic importance (section 17), journalistic content (section 19) and fraudulent advertising content (Chapter 5 of Part 3) duties will be considered in Codes for Category 1 services. Content reporting and complaints procedures, as regards these types of content specifically, may also feature in these Codes.

A1.5 These Codes of Practice:

a) relate to the design, operation and use of a user-to-user service—
   i) in the United Kingdom, or
   ii) as they affect United Kingdom users of the service; and
b) apply regardless of whether or not the person who provides the service is inside the United Kingdom.

A1.6 So far as relating to a user-to-user service that includes regulated provider pornographic content, these Codes do not extend to—

a) the regulated provider pornographic content, or
b) the design, operation or use of the service so far as relating to that content.

A1.7 So far as relating to a user-to-user service that is a combined service, these Codes do not extend 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.

A1.8 These Codes come into force on [in our final Codes, we will specify a day that is 21 days from the publication date].

Illegal content Codes of Practice

A1.9 Recommended measures for search services are set out separately in the accompanying Codes of Practice for search services.

A1.10 While a small number of recommended measures apply to providers in relation to all relevant regulated services, including those provided by small and microbusinesses, the majority of our recommended measures apply only in relation to services that have identified certain risks or are of a certain size. The application of each recommendation is specified within the relevant measure, and an overview can be found in the **Index of recommended measures** in Chapter 2. The section headed 'Risk and risk of harm' (which begins at paragraph A11.4) explains how to interpret provisions relating to being 'at risk of' kinds of harm, and a definition of 'multi-risk' is included in the definitions section at the end of the Codes. The section headed 'User numbers' (which begins at paragraph A11.6) explains how to count users for those measures which apply in relation to services of a certain size, and the definition of 'large service' is included in the definitions section at the end of the Codes.

A1.11 Ofcom is required to develop separate Codes for terrorism (arising from the offences set out in Schedule 5 to the Act), child sexual exploitation and abuse ('**CSEA**') (arising from the offences set out in Schedule 6 to the Act), and one or more Codes for the purpose of compliance with the relevant duties relating to illegal content and harms (except to the extent measures are included in the Codes for terrorism and CSEA). Many of our recommended measures apply to more than one illegal harm. To minimise duplication and simplify the regime for service providers, we have produced one document containing the Codes for terrorism, CSEA and other duties. We identify the relevant Code(s) for each measure in the **Index of recommended measures.**

Enforcement of these Codes

A1.12 The recommended measures in these Codes apply to providers in respect of the regulated user-to-user service that they provide. If a person is the provider of more than one regulated user-to-user service, the recommended measures in these Codes have effect in relation to each such service (so far as applicable). Each recommendation is accompanied by an 'application' section specifying the services in respect of which it applies.

A1.13 The Act provides that service providers which implement measures recommended to them in these Codes will be treated as complying with the relevant duty or duties to which those measures relate. We have mapped our recommended measures against the relevant duties in the Index of recommended measures.

A1.14 Where a service provider implements measures recommended to it in these Codes which include safeguards for the protection of freedom of expression and/or for the protection of users' privacy, the Act provides that they will also be treated as complying with the duties set out in section 22(2) (in respect of freedom of expression) and section 22(3) (in respect of privacy).

A1.15 Service providers may seek to comply with a relevant duty in another way by adopting what the Act refers to as alternative measures. If doing so, service providers would also need to comply with the duty to have particular regard to the importance of protecting users' rights to freedom of expression within the law, and of protecting the privacy of users.

A1.16 Where they take alternative measures, service providers must also maintain a record of what they have done and how they consider that it meets the relevant duties, including how they have complied with the duty to have particular regard to the importance of protecting freedom of expression and privacy.

A1.17 Alongside these Codes, please refer to our separate guidance on:

a) ***Risk assessment***: some of the measures in these Codes apply where a service provider has, through its risk assessment, identified a high- or medium-risk of harm on the service. We expect all service providers to conduct a suitable and sufficient risk assessment, referring to our guidance.
b) ***Illegal content judgements***: service providers will need to understand what illegal content is, under the Act.
c) ***Record-keeping***: under the Act, service providers are required to keep records of (1) steps that they have taken in accordance with these Codes, or (2) any alternative steps they are taking to comply with their duties.
d) ***Enforcement***: to find out more about Ofcom's approach to enforcement, please refer to our Enforcement guidance.

Other obligations on regulated service providers

A1.18 The recommendations in these Codes do not affect other regulatory and legislative requirements on providers of services regulated under the Act. They will also need to ensure that they comply with data protection law and, where relevant, the Privacy and Electronic Communications (EC Directive) Regulations 2003 (PECR). Users' rights to data protection are covered by UK GDPR and the Data Protection Act 2018 which are enforced by the Information Commissioner's Office (the '**ICO**'). The ICO has a range of data protection and PECR compliance guidance which services may wish to consult. Services likely to be accessed by children should also ensure they conform with the ICO's Children's code.

A2. Using these Codes

Structure of the Codes

A2.1 Chapters of the Codes refer to the thematic area of the recommended measures – for example, if they relate to governance and accountability, or content moderation functions.

A2.2 The application of a recommendation is set out in the section under the subheading 'Application' in each measure.

A2.3 Definitions of terms in **bold** are set out in Chapter 11 (definitions and interpretation).

A2.4 We have also provided a table below indexing the recommended measures according to:

a) the services in relation to which they apply;
b) the Codes in which they are included; and
c) the duties to which they relate.

Index of recommended measures and which Code(s) they are in

Chapter	Recommended measure	Application	Relevant Codes	Relevant Duties
Governance and accountability	3A Annual review of risk management activities	**Large services** only.	• CSEA • Terrorism • Other duties	Section 10, 20, 21
	3B Person accountable for illegal content safety duties and reporting and complaints duties	All **services.**		Section 10, 20, 21
	3C Written statements of responsibilities	**Large** or **multi-risk services**.		Section 10, 20, 21
	3D Internal monitoring and assurance	**Large services** that are **multi-risk services.**		Section 10, 20, 21

Chapter	Recommended measure	Application	Relevant Codes	Relevant Duties
	3E Tracking evidence of new and increasing illegal harm	**Large** or **multi-risk services**.		Section 10(2), 10(3)
	3F Code of conduct regarding protection of users from illegal harm			Section 10, 20, 21
	3G Staff compliance training			Section 10, 20, 21
Content moderation	4A Having a content moderation function that allows for the swift take down of illegal content	All **services.**	• CSEA • Terrorism • Other duties	Section 10(2), 10(3), 10(4)(e), 10(6)
	4B Setting internal content policies	**Large** or **multi-risk services.**		Section 10(2), 10(3), 10 (4)(e), 10(6)
	4C Performance targets			Section 10(2), 10(3), 10(4)(e), 10(6)
	4D Prioritisation			Section 10(2), 10(3), 10(4)(e)
	4E Resourcing			Section 10(2), 10(3), 10(4)(e), 10(4(h)), 10(6)
	4F Provision of training and materials to moderators			Section 10(2), 10(3), 10(4)(e), 10(4)(h), 10(6)
	4G Hash matching for CSAM	**Large services** that are at medium or high **risk** of image-based CSAM OR services that are at high **risk** of image-based CSAM and (a) have more than 700,000 monthly UK users or (b) are **file-storage and file-sharing services** and have more than 70,000 monthly UK users.	• CSEA	Section 10(2), 10(3), 10(4)(e)

Chapter	Recommended measure	Application	Relevant Codes	Relevant Duties
	4H Detection of CSAM URLs	**Large services** that are at medium or high **risk** of CSAM URLs OR services that have more than 700,000 monthly UK users and are at high risk of CSAM URLs.		Section 10(2), 10(3), 10(4)(e)
	4I Keyword detection regarding articles for use in frauds	**Large services** that are at medium or high **risk** of fraud.	• Other duties	Section 10(2), 10(3), 10(4)(e)
Reporting and complaints	5A Enabling complaints	All **services.**	• CSEA • Terrorism • Other duties	Section 21(2)(a)
	5B Having an easy to find, easy to access and easy to use complaints system			Section 20, 21
	5C Appropriate action - sending indicative timelines			Section 21
	5D Appropriate action for relevant complaints about suspected illegal content			Section 21
	5E Appropriate action for relevant complaints which are appeals - determination	(i) **Large** or **multi-risk services** and, separately, (ii) all other **services**.		Section 21
	5F Appropriate action for relevant complaints which are appeals – action following determination	All **services.**		Section 21
	5G Appropriate action for relevant complaints about proactive technology, which are not appeals			Section 21

Chapter	Recommended measure	Application	Relevant Codes	Relevant Duties
	5H Appropriate action for all other relevant complaints			Section 21
	5I Dedicated reporting channels	**Large services** that are at medium or high **risk** of fraud.	• Other duties	Section 10(3)
Terms of service	6A Substance of the terms	All **services**	• CSEA • Terrorism • Other duties	Section 10(4)(c) 10(5), 10(7), 21(3)
	6B Clarity and accessibility			Section 10(8), 21(3)
Default settings and support for child users	7A Safety defaults for child users	All services that are at high **risk** of grooming OR **large services** at medium **risk** of grooming, and in each case, which have an existing means of identifying **child users**.	• CSEA • Other duties	Section 10(2), 10(4)(d), 10(4)(f)
	7B Support for child users		• CSEA	Section 10(2), 10(4)(g)
Recommender systems	8A Safety metrics for on-platform testing of recommender systems	Services that conduct **on-platform testing** of **recommender systems** and are at medium or high **risk** of at least two specified harms.	• CSEA • Terrorism • Other duties	Section 10(2), 10(4)(b)
Enhanced user controls	9A User blocking and muting	**Large services** that are at medium or high **risk** of one or more specified harms, have **user profiles** and have at least one specified functionality.	• CSEA • Other duties	Section 10(2), 10(4)(f)
	9B Disabling comments	**Large services** that are at medium or high **risk** of one or more specified harms and enable **users** to comment on content.	• CSEA • Other duties	Section 10(2), 10(4)(f)

Chapter	Recommended measure	Application	Relevant Codes	Relevant Duties
	9C User verification/labelling schemes	**Large services** that are at medium or high **risk** of one or more specified harms, which label the **user profiles** to indicate the account is participating in a **notable user scheme** or subscribed to a **monetised scheme**.	• Other duties	Section 10(2), 10(4)(b)
User access	10A Removing accounts of proscribed organisations	All **services.**	• Terrorism	Section 10(2), 10(3), 10(4)(d)

A3. Governance and Accountability

3A. Annual review of risk management activities

Application

A3.1 This measure applies to a **provider** in respect of each **large service** it provides.

Recommendation

A3.2 The **provider**'s most senior **governance body** in relation to the **service** should carry out and record an annual review of risk management activities in relation to online safety, and how developing risks are being monitored and managed.

3B. Person accountable for illegal content safety duties and reporting and complaints duties

Application

A3.3 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A3.4 The **provider** should name a person accountable to the most senior **governance body** for compliance with the **illegal content safety duties** and the **reporting and complaints duties**.

A3.5 Being accountable means being required to explain and justify actions or decisions regarding online safety risk management and mitigation, and compliance with the relevant duties, to the most senior **governance body**.

3C. Written statements of responsibilities

Application

A3.6 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A3.7 The **provider** should have written statements of responsibilities for senior members of staff who make decisions related to the management of online safety risks.

A3.8 A statement of responsibilities is a document which clearly shows the responsibilities that the senior manager performs in relation to online safety risk management and how they fit

in with the **provider**'s overall governance and management arrangements in relation to the **service**.

3D. Internal monitoring and assurance

Application

A3.9 This measure applies to a **provider** in respect of each **service** it provides that is both a **large service** and a **multi-risk service**.

Recommendation

A3.10 The **provider** should have an internal monitoring and assurance function to provide independent assurance that measures taken to mitigate and manage the risks of harm to individuals identified in the **risk assessment** are effective on an ongoing basis, reporting to either:

a) the body that is responsible for overall governance and strategic direction of a service; or
b) an audit committee.

A3.11 This independent assurance may be provided by an existing internal audit function.

3E. Tracking evidence of new and increasing illegal harm

Application

A3.12 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A3.13 The **provider** should track evidence of new kinds of **illegal content** on the **service**, and unusual increases in particular kinds of **illegal content** or **illegal content proxy**, or equivalent changes in the use of the **service** for the commission or facilitation of **priority offences**. Relevant evidence may include, but is not limited to, that derived from:

a) complaints processes;
b) content moderation processes;
c) referrals from law enforcement; and
d) information from **trusted flaggers** and any other expert group or body the **provider** considers appropriate.

A3.14 The **provider** should regularly report any new kinds of **illegal content** or unusual increases in particular kinds of **illegal content** or **illegal content proxy**, or equivalent changes in the use of the **service** for the commission of facilitation of **priority offences** through relevant governance channels to the most senior **governance body**.

A3.15 To understand this, the **provider** should establish a baseline understanding of how frequently particular kinds of **illegal content**, **illegal content proxy**, or the commission or facilitation of **priority offences** occur on the service to the extent possible based on its internal data and evidence. The **provider** should use this baseline to identify unusual increases in the relevant data.

3F. Code of conduct regarding protection of users from illegal harm

Application

A3.16 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A3.17 The **provider** should have a Code of Conduct that sets standards and expectations for employees around protecting **users** from risks of **illegal harm**.

3G. Staff compliance training

Application

A3.18 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A3.19 The **provider** should secure that staff involved in the design and operational management of the **service** are trained in the **service**'s approach to compliance with the **illegal content safety duties** and the **reporting and complaints duties**, sufficiently to give effect to it.

A3.20 This does not affect Recommendations 4F (training and materials) and 9C (user verification/labelling schemes) (see paragraph A9.12(f)).

A4. Content moderation

4A. Having a content moderation function that allows for the swift take down of illegal content

Application

A4.1 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A4.2 The **provider** should have systems or processes designed to swiftly take down **illegal content** of which it is aware (a '**content moderation function**').

A4.3 For this purpose, when the **provider** has reason to suspect that content may be **illegal content**, the **provider** should either:

a) make an illegal content judgement in relation to the **content** and, if it determines that the **content** is **illegal content**, swiftly take the **content** down; or
b) where the **provider** is satisfied that its **terms of service** prohibit the types of **illegal content** which it has reason to suspect exist, consider whether the **content** is in breach of those **terms of service** and, if it is, swiftly take the **content** down.

A4.4 This does not affect Recommendations 4G (hash matching for CSAM) and 4H (detection of CSAM URLs) (see paragraphs A4.24 and A4.38 respectively).

Safeguards for freedom of expression

A4.5 The following measures, where applicable, are safeguards to protect users' right to freedom of expression:

a) Recommendation 4C (performance targets), so far as it relates to the accuracy of decision making;
b) Recommendation 4F (training and materials); and
c) Recommendations 5E(i), 5E(ii) and 5F (appeals).

4B. Setting internal content policies

Application

A4.6 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A4.7 The **provider** should set and record (but need not necessarily publish) internal content policies setting out rules, standards and guidelines around:

a) what **content** is allowed on the **service** and what is not; and
b) how policies should be operationalised and enforced.

A4.8 The policies should be drafted in such a way that **illegal content** (where it is identifiable as such) is not permitted.

A4.9 In setting such policies, the **provider** should have regard to:

a) the **risk assessment** of the **service**; and
b) information pertaining to the tracking of signals of emerging **illegal harm**.

4C. Performance targets

Application

A4.10 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A4.11 The **provider** should set and record performance targets for its **content moderation function**, covering at least:

a) the time that **illegal content** remains on the **service** before it is taken down; and
b) the accuracy of decision making.

A4.12 In setting its targets, the **provider** should balance the desirability of taking **illegal content** down swiftly against the desirability of making accurate moderation decisions.

A4.13 The **provider** should effectively measure and monitor its performance against its performance targets.

4D. Prioritisation

Application

A4.14 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A4.15 The **provider** should prepare and apply a policy in respect of the prioritisation of **content** for review. In setting the policy, the **provider** should have regard to at least the following:

a) the virality of **content**: the **provider** should prioritise **content** for review in a way which minimises circumstances in which the number of **users encountering** a particular item of **illegal content** increases exponentially over a period of time;

b) the potential severity of **content**: including whether the **content** is suspected to be **priority illegal content** and the **risk assessment** of the service; and
c) the likelihood that **content** is **illegal content**, including whether it has been reported by a **trusted flagger**.

4E. Resourcing

Application

A4.16 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A4.17 The **provider** should resource its **content moderation function** so as to give effect to its internal content policies and performance targets having regard to at least:

a) the propensity for external events to lead to a significant increase in demand for content moderation on the **service**; and
b) the particular needs of its **United Kingdom user** base as identified in its **risk assessment**, in relation to languages.

4F. Provision of training and materials to moderators

Application

A4.18 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A4.19 The **provider** should ensure people working in content moderation receive training and materials that enable them to moderate **content** in accordance with Recommendations 4A and 4B. This measure does not apply in relation to **volunteers**.

A4.20 The **provider** should ensure that in doing so:

a) it has regard to at least the **risk assessment** of the **service** and information pertaining to the tracking of signals of emerging **illegal harm**; and
b) where the **provider** identifies a gap in moderators' understanding of a specific kind of **illegal harm**, it gives training and materials to remedy this.

4G. Hash matching for CSAM

Application

A4.21 This measure applies to a **provider** in respect of each **service** it provides that:

a) is at high **risk** of image-based CSAM, and:

i) has more than 700,000 **monthly United Kingdom users**; or
ii) is a file-storage and file-sharing service and has more than 70,000 monthly United Kingdom users; or

b) is a **large service** and is at medium or high **risk** of image-based CSAM.

Key definition

A4.22 In this Recommendation "relevant content" means:

a) any **regulated user-generated content** in the form of photographs, videos or visual material that:

i) may be **encountered** by **United Kingdom users** of the service by means of the service, and
ii) is communicated publicly[1] by means of the service; or

b) any material which, if it were present on the service, would be content within sub-paragraph (a).

Recommendation

A4.23 The **provider** should ensure that, where technically feasible**, perceptual hash matching technology** is used effectively to analyse relevant content to assess whether it is **CSAM**.

A4.24 The provider should ensure that appropriate measures are taken to swiftly **take down** (or prevent from being generated, uploaded or shared) **detected content** that is **CSAM** (but see paragraphs A4.30 to A4.33 in relation to the use of human moderators).

A4.25 For the use of the **perceptual hash matching technology** to be effective, it should:

a) analyse all relevant content present on the service at the time the technology is implemented within a reasonable time;
b) analyse relevant content that is generated on, uploaded to or shared on the service (or that a user seeks to so generate, upload or share) after the technology has been implemented before or as soon as practicable after it can be **encountered** by **United Kingdom users** of the service;
c) use a suitable perceptual hash function to compare relevant content to the latest version of an appropriate hash database (see paragraph A4.26); and
d) be configured so that its performance strikes an appropriate balance between **precision** and **recall** (see paragraphs A4.27 to A4.29).

A4.26 For a hash database to be appropriate, the following conditions should be met:

[1] Ofcom has published [*draft*] **guidance on content communicated "publicly" and "privately" under the Online Safety Act** for this purpose. (This footnote is for information and does not form part of the Code of Practice.)

a) the hash database should include hashes of **CSAM** sourced from a person with expertise in the identification of **CSAM**, and who has arrangements in place to identify suspected **CSAM**;
b) the provider should ensure that arrangements are in place in relation to the hash database (including any hashes sourced from a person as mentioned in sub-paragraph (a)) to:

i) secure (so far as possible) that **CSAM** is correctly identified before hashes of that material are added to the database (such as assessment by persons with expertise in making such judgements);
ii) regularly update the database with hashes of **CSAM**;
iii) review cases where material is suspected to have been incorrectly identified as **CSAM**, and remove such hashes from the database where appropriate; and

c) the **provider** should ensure that appropriate measures are taken to secure the hash database from unauthorised access, interference or exploitation (whether by persons who work for the **provider** or are providing a service to the **provider**, or any other person).

A4.27 In configuring the technology so that its performance strikes an appropriate balance between **precision** and **recall**, the provider should ensure that the following matters are taken into account:

a) the risk of harm relating to image-based **CSAM**, as identified in the **risk assessment** of the service, and including in particular information reasonably available to the **provider** about the prevalence of relevant content that is **CSAM** on the service;
b) the proportion of **detected content** that is a **false positive**; and
c) the effectiveness of the systems and/or processes used to identify **false positives**.

A4.28 The **provider** should ensure that the performance of the technology, and whether the balance between **precision** and **recall** continues to be appropriate, is reviewed at least every six months.

A4.29 The **provider** should ensure that a written record is made of how this balance has been struck in configuring the technology, including what information has been considered, and information about reviews and steps taken in response.

A4.30 The **provider** should ensure that human moderators are used to review an appropriate proportion of content **detected** as **CSAM**.

A4.31 When deciding what proportion of **detected content** it is appropriate to review, the **provider** should ensure that the following principles are taken into account:

a) the resource dedicated to review of **detected content** should be proportionate to the degree of accuracy achieved by the **perceptual hash matching technology** in use and any associated systems and/or processes (as indicated by the periodic reviews of the performance of the technology mentioned in paragraph A4.28, and taking account of the outcomes of reviews of content carried out by human moderators and data from the complaint procedure enabling **users** to complain if they believe their content has wrongly been identified as **illegal content**);
b) this resource should be targeted at content with a higher likelihood of being a **false positive**.

A4.32 The **provider** should ensure that a written record is made of its policy for review of **detected content**, which sets out:

a) the proportion of **detected content** which is intended to be reviewed; and
b) information about how the principles mentioned in paragraph A4.31 were taken into account in setting that policy.

A4.33 The **provider** should keep statistical records about content reviewed in accordance with that policy (including the number of reviews carried out, the proportion of **detected content** that this represents, and the number of **false positives** identified).

Safeguards for freedom of expression and privacy

A4.34 Sub-paragraphs (c) and (d) of paragraph A4.25 and paragraphs A4.26 to A4.33 are safeguards to protect users' right to freedom of expression and privacy.

4H. Detection of CSAM URLs

Application

A4.35 This measure applies to a **provider** in respect of each **service** it provides that:

a) has more than 700,000 **monthly United Kingdom users** and is at high **risk** of CSAM URLs; or
b) is a **large service** and is at medium or high **risk** of CSAM URLs.

Key definitions

A4.36 In this Recommendation:

"CSAM URL" means 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**);

"Relevant content" means:

a) any **regulated user-generated content** in the form of written material or messages (including hyperlinks) that:
   i) may be **encountered** by **United Kingdom users** of the service by means of the service, and
   ii) is communicated publicly[2] by means of the service; or
b) any material which, if it were present on the service, would be content within sub-paragraph (a).

[2] Ofcom has published [*draft*] **guidance on content communicated "publicly" and "privately" under the Online Safety Act** for this purpose. (This footnote is for information and does not form part of the Code of Practice.)

Recommendation

A4.37 The **provider** should ensure that, where technically feasible, technology is used effectively to analyse relevant content to assess whether it consists of or includes a CSAM URL.

A4.38 The provider should ensure that content **detected** to be a CSAM URL is swiftly **taken down** (or prevented from being generated, uploaded or shared).

A4.39 For the use of the technology to be effective, it should:

a) analyse all relevant content present on the service at the time the technology is implemented within a reasonable time;
b) analyse relevant content that is generated on, uploaded to or shared on the service (or that a user seeks to so generate, upload or share) after the technology has been implemented before or as soon as practicable after it can be **encountered** by **United Kingdom users** of the service;
c) compare analysed content to the latest version available of an appropriate list (see paragraph A4.40); and
d) **detect** content to be a CSAM URL where it is a **URL** that directly matches a **URL** on the list (except that it does not matter whether or not an access protocol, such as "https://", is included) or that contains a domain on the list.

A4.40 For a list to be appropriate, the **provider** should source it from a person with expertise in the identification of **CSAM**, and who has arrangements in place to:

a) identify suspected CSAM URLs;
b) secure (so far as possible) that **URLs** at which **CSAM** is present, and domains which are entirely or predominantly dedicated to **CSAM**, are correctly identified before they are added to the list;
c) regularly update the list with identified CSAM URLs;
d) review CSAM URLs on the list, and remove any which are no longer CSAM URLs; and
e) secure the list from unauthorised access, interference or exploitation (whether by persons who work for that person, or by any other person).

A4.41 The **provider** should ensure that appropriate measures are taken to secure any copy of the list held for the purposes of this Recommendation from unauthorised access, interference or exploitation (whether by persons who work for the **provider** or are providing a service to the **provider**, or any other person).

Safeguards for freedom of expression and privacy

A4.42 The arrangements specified in sub-paragraphs (b), (d) and (e) of paragraph A4.40, and the measures mentioned in paragraph A4.41, are safeguards to protect users' right to freedom of expression and privacy.

4I. Keyword detection regarding articles for use in frauds

Application

A4.43 This measure applies to a **provider** in respect of each **service** it provides that is both a **large service** and at medium or high **risk** of fraud.

Key definition

A4.44 In this Recommendation "relevant content" means:

a) any **regulated user-generated content** in the form of written material or messages that:

i) may be **encountered** by **United Kingdom users** of the service by means of the service, and

ii) is communicated publicly[3] by means of the service; or

b) any material which, if it were present on the service, would be content within sub-paragraph (a).

Recommendation

A4.45 The **provider** should ensure that, where technically feasible, **fuzzy keyword detection technology** is used effectively to analyse relevant content to assess whether it is likely to amount to an **offence concerning articles for use in frauds**.

A4.46 The **provider** should ensure that **detected content** is considered in accordance with the service's internal content policies (see Recommendation 4B (setting internal content policies)).

A4.47 For the use of the technology to be effective, it should:

a) analyse all relevant content present on the service at the time the technology is implemented within a reasonable time;
b) analyse relevant content that is generated on, uploaded to or shared on the service (or that a user seeks to so generate, upload or share) after the technology has been implemented before or as soon as practicable after it can be **encountered** by **United Kingdom users** of the service;
c) use a suitable keyword list (see paragraph A4.48); and
d) be configured so that its performance strikes an appropriate balance between **precision** and **recall** (see paragraphs A4.53 to A4.55).

A4.48 For a keyword list to be suitable, the **provider** should ensure that appropriate steps are taken to ensure that it:

a) contains only **keywords** that could not reasonably be expected to be used on the service (either on their own or in combination with other **keywords** on the keyword list) except in relation to the commission of an **offence concerning articles for use in frauds**; and
b) is sufficiently comprehensive.

A4.49 The steps referred to in paragraph A4.48 include:

a) The compilation of an initial list of **keywords** that are (either on their own or in combination with other **keywords** on the list) unlikely to be used except in relation to the commission of an **offence concerning articles for use in frauds**, taking account of:

i) 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

[3] Ofcom has published [*draft*] **guidance on content communicated "publicly" and "privately" under the Online Safety Act** for this purpose. (This footnote is for information and does not form part of the Code of Practice.)

ii) the outcomes of reviews of **content** carried out by human moderators, **reports**, and **content** flagged or reported by trusted flaggers or similar systems;

b) the taking of appropriate measures to test the use of those **keywords** (either on their own or in combination, as appropriate) on a reasonable sample of relevant content, and review any content detected by that testing to identify **false positives**. What is a reasonable sample size will depend on the volume of relevant content present on the service;

c) the removal of any **keywords** from the initial list which, in light of the measures taken in accordance with sub-paragraph (b), cannot reasonably be expected to be used (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**;

d) the taking of appropriate measures to secure the keyword list from unauthorised access, interference or exploitation (whether by persons who work for the **provider** or are providing a service to the **provider**, or any other person); and

e) the review of the keyword list in accordance with paragraphs A4.50 to A4.51.

A4.50 Subject to paragraph A4.51, the **provider** should ensure that the keyword list is reviewed at least every six months, taking account of:

a) 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 **provider**; and

b) evidence reasonably available to the **provider** on the accuracy and effectiveness of the **fuzzy keyword detection technology** in detecting relevant content that amounts to an **offence concerning articles for use in frauds**. This includes evidence from the outcomes of reviews of **content** carried out by human moderators, **reports**, and **content** flagged or reported by trusted flaggers or similar systems.

A4.51 The **provider** should ensure that the keyword list is reviewed more frequently than every six months where evidence reasonably available to it suggests that this would be proportionate. Such evidence includes that reviews of **detected content** by human moderators are identifying a large volume of **false positives**, or evidence that a large volume of relevant content that amounts to an **offence concerning articles for use in frauds** is not being detected by use of the technology.

A4.52 The **provider** should ensure that a written record is made of:

a) the steps that it took to compile its keyword list (including what information it considered and what measures it took in accordance with paragraph A4.49(b)), and what **keywords** it included in that list; and

b) each review of the keyword list, including the date of that review, what information it considered, the measures it took in accordance with paragraph A4.49(b), and any steps taken in response to the review (including any modifications to that list).

A4.53 In configuring the technology so that its performance strikes an appropriate balance between **precision** and **recall**, the **provider** should ensure that the following matters are taken into account:

a) the risk of harm relating to **fraud**, as identified in the **risk assessment** of the **service**, and including in particular information reasonably available to the **provider** about the prevalence of relevant content that amounts to an **offence concerning articles for use in frauds** on the service;

b) the proportion of **detected content** that is a **false positive**; and
c) the effectiveness of any systems and/or processes used to identify **false positives** before **detected content** is **taken down**.

A4.54 The **provider** should ensure that the performance of the technology, and whether the balance between **precision** and **recall** continues to be appropriate, is reviewed at the same time as the review of the keyword list.

A4.55 The **provider** should ensure that a written record is made of how this balance has been struck in configuring the technology, including what information has been considered, and information about reviews and steps taken in response.

A4.56 The **provider** should ensure that human moderators are used to review a reasonable sample of **detected content** within each period between reviews to identify **false positives**. When deciding what is a reasonable sample of **detected content** to review, the **provider** should ensure the following principles are taken into account:

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

i) the volume of **detected content** in that period;
ii) the volume of **false positives** identified by human moderators in the preceding period; and
iii) data from the complaint procedure enabling **users** to complain if they believe their **content** has wrongly been identified as **illegal content** in that period.

b) 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**.

A4.57 The provider should ensure that a written record is made in respect of each period between reviews which sets out:

a) the volume of **detected content** reviewed by human moderators within that period;
b) the proportion of such content that is a **false positive**; and
c) information about how the principles mentioned in paragraph A4.56 were taken into account.

Safeguards for freedom of expression and privacy

A4.58 Sub-paragraphs (c) and (d) of paragraph A4.47, sub-paragraph (a) of paragraph A4.48 and paragraphs A4.49 to A4.57 are safeguards to protect users' right to freedom of expression and privacy.

A4.59

A5. Reporting and complaints

5A. Enabling complaints

Application

A5.1 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A5.2 The **provider** should have complaints processes which enable **United Kingdom users** and **affected persons** to make each type of **relevant complaint** in a way which will secure that the **provider** will take appropriate action in relation to them.

5B. Having an easy to find, easy to access and easy to use complaints system

Application

A5.3 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A5.4 The **provider's** systems and processes for making **relevant complaints** should be operated to ensure that:

a) for **relevant complaints** regarding a specific piece of **content**, a reporting function or tool is clearly accessible in relation to that **content**;
b) processes for making other kinds of **relevant complaints** should be easy to find and easily accessible;
c) the number of steps necessary (such as the number of clicks or navigation points) to submit (i) a **relevant complaint** using the reporting function or tool; and (ii) any other kind of **relevant complaint** are as few as is reasonably practicable; and
d) **users** and **affected persons** have the ability when making **relevant complaints** to provide the **provider** with relevant information or supporting material.

A5.5 In designing its complaints processes for **relevant complaints**, including its reporting tool or function, the **provider** should have regard to the particular needs of its **United Kingdom user** base as identified in its **risk assessment**. This should include the particular needs of:

a) children (for services likely to be accessed by children and considering the likely age of the children using that service); and
b) disabled people.

A5.6 For the purposes of paragraph A5.5(a), any written information for **users** comprised in the system or process should be comprehensible based on the likely reading age of the youngest person permitted to agree to the **service's terms of service**.

A5.7 For the purposes of paragraph A5.5(b), the system or process should be designed for the purposes of ensuring usability for those dependent on assistive technologies including:

a) keyboard navigation; and
b) screen reading technology

5C. Appropriate action – sending indicative timelines

Application

A5.8 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A5.9 The **provider** should acknowledge receipt of each **relevant complaint** and provide the complainant with an indicative timeframe for deciding the complaint.

5D. Appropriate action for relevant complaints about suspected illegal content

Application

A5.10 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A5.11 When the provider receives a relevant complaint about suspected illegal content:

a) if the **provider** has established a process for **content** prioritisation and applicable performance targets, it should handle the complaint in accordance with them;
b) if the **service** has no process for **content** prioritisation and applicable performance targets it should consider the complaint promptly; and
c) in either case, it should act in accordance with Recommendation 4A (having a **content moderation function**) in relation to the suspected **illegal content**.

5E(i). Appropriate action for relevant complaints which are appeals – determination (large or multi-risk services)

Application

A5.12 This measure applies to a **provider** in respect of each **service** it provides that is either (or both) of the following:

a) a **large service**; or
b) a **multi-risk service**.

Recommendation

A5.13 For the determination of **relevant complaints** which are **appeals**, the **provider** should set, and monitor its performance against, performance targets relating to at least the time it

takes to determine the **appeal** and the accuracy of decision making, and should resource itself so as to give effect to those targets.

A5.14 The **provider** should have regard to the following matters in determining what priority to give to review of a **relevant complaint** which is an **appeal**:

a) the severity of the action taken against the **user** as a result of the decision that the **content** was **illegal content**;
b) whether the decision that the **content** was **illegal content** was made by **proactive technology** and the likelihood of **false positives** generated by the specific **proactive technology** used; and
c) the **service's** past error rate in making illegal content judgements of the type concerned.

5E(ii). Appropriate action for relevant complaints which are appeals – determination (services that are neither large nor multi-risk)

Application

A5.15 This measure applies to a **provider** in respect of each **service** it provides that is neither a **large service** nor a **multi-risk service**.

Recommendation

A5.16 The **provider** should determine **relevant complaints** which are **appeals** promptly.

5F. Appropriate action for relevant complaints which are appeals – action following determination

Application

A5.17 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A5.18 If, in relation to a **relevant complaint** that is an **appeal**, the **provider** reverses a decision that **content** was **illegal content**, the **provider** should:

a) restore the **content** and/or the **user's user account** to the position they would have been in had the **content** not been judged to be **illegal content**;
b) where necessary to avoid similar errors in future, adjust the relevant content moderation guidance; and
c) where necessary to avoid similar errors in future, take such steps as are within its power to secure that the use of automated content moderation technology does not cause the same **content** to be taken down again.

5G. Appropriate action for relevant complaints about proactive technology, which are not appeals

Application

A5.19 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A5.20 This measure relates to **relevant complaints**, which are not **appeals**, about the use of **proactive technology** on the **service** when:

a) the use of **proactive technology** on the **service** results in **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
b) 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 blocking **content** not of a kind specified in the **terms of service** as a kind of **content** in relation to which the technology would operate).

A5.21 The **provider** should inform the complainant of their right, if they consider the **provider** to be in breach of contract, to bring proceedings.

5H. Appropriate action for all other relevant complaints

Application

A5.22 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A5.23 The **provider** should establish a triage process for **relevant complaints** with a view to protecting **users** from harm, including harm to their rights. A responsible person, team or function should be nominated to lead this triage process and ensure **relevant complaints** reach the most relevant function or team.

A5.24 **Relevant complaints** should be dealt with:

a) in a way that protects **users** and the **provider's** compliance with other applicable laws in question;
b) within timeframes the **provider** has determined are appropriate; and
c) in accordance with Recommendations 5D to 5G.

5I. Dedicated reporting channels

Application

A5.25 This measure applies to a **provider** in respect of each **service** it provides that is both a **large service** and at medium or high **risk** of fraud.

Recommendation

A5.26 The **provider** should establish and maintain a dedicated reporting channel in the circumstances set out in this Recommendation.

A5.27 In this Recommendation, a 'trusted flagger' is each of the following:

a) HM Revenue and Customs (HMRC);
b) Department for Work and Pensions (DWP);
c) City of London Police (CoLP);
d) National Crime Agency (NCA);
e) National Cyber Security Centre (NCSC);
f) Dedicated Card Payment Crime Unit (DCPCU);
g) Financial Conduct Authority (FCA).

A5.28 The **provider** should publish a clear and accessible policy on its processes relating to the establishment of a dedicated reporting channel for **trusted flaggers**, covering any relevant procedural matters. The policy should include a commitment from the **provider** to engage with a **trusted flagger** to understand its needs with respect to the dedicated reporting channel.

A5.29 If a request is made in accordance with the policy by a **trusted flagger**, the **provider** should establish and maintain a dedicated reporting channel for fraud.

A5.30 At least every two years, the provider should seek feedback from the **trusted flaggers** with which it has made such arrangements, on whether any reasonable adjustments or improvements might be made to the operation of the dedicated reporting channel.

A5.31 Complaints from **trusted flaggers** received through the dedicated reporting channel relating to specific **content** should be handled in accordance with Recommendations 4A to 4E (content moderation). **Providers** should ensure that complaints received through the dedicated reporting channel relating to other matters are handled as if they were **relevant complaints**, in accordance with Recommendation 5G (appropriate action for all other relevant complaints) and, where relevant, Recommendation 3E (tracking evidence of new and increasing illegal harm).

A6. Terms of service

6A. Substance of the terms

Application

A6.1 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A6.2 The **provider** should include the following in its **terms of service**:

a) provisions specifying how individuals are to be protected from **illegal content**, addressing:

i) separately for each of **terrorism content**, **CSEA content** and other **priority illegal content**, how the provider will minimise the length of time for which any **priority illegal content** is present; and

ii) how, where the **provider** is alerted by a person to the presence of any **illegal content**, or becomes aware of it in any other way, it will swiftly **take down** such **content**.

b) provisions giving information about any **proactive technology** used for the purposes of compliance with any of the **illegal content safety duties**[4] (including the kind of technology, when it is used, and how it works);

c) provisions specifying the policies and processes that govern the handling and resolution of **relevant complaints**.

6B. Clarity and accessibility

Application

A6.3 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A6.4 The **provider** should ensure that the provisions included in **terms of service** in accordance with Recommendation 6A are:

a) easy to find, such that they are:

i) clearly signposted for the general public regardless of whether they have signed up to or are using the **service**; and

ii) locatable within the terms of **service**;

b) laid out and formatted in a way that helps **users** read and understand them;

c) written to a reading age comprehensible for the youngest person permitted to agree to them; and

[4] For this purpose, "illegal content safety duties" means the duties in section 10(2) and (3) of the Online Safety Act 2023.

d) designed for the purposes of ensuring usability for those dependent on assistive technologies, including:

i) keyboard navigation; and
ii) screen reading technology.

A6.5

A7. Default settings and user support for child users

7A. Safety defaults for child users

Application

A7.1 This measure applies to a **provider** in respect of each **service** it provides that is either of the following, to the extent that it has an existing means of identifying **child users** of the **service** concerned:

a) at high **risk** of grooming; or
b) a **large service** that is at medium **risk** of grooming.

Recommendation

A7.2 If the service has network expansion prompts or connection lists, the provider should implement default settings ensuring that:

a) **child users** are not included in **network expansion prompts** presented to other **users**;
b) **child users** are not presented with **network expansion prompts**;
c) **child users** are not included in the **connection lists** of other **users**;
d) **connection lists** of **child users** are not displayed to other **users**;

A7.3 If the service has **direct messaging**, the **provider** should implement **default settings** ensuring that:

a) If the service has user connections, child users cannot receive direct messages from a non-connected user;
b) If the **service** does not have **user connections**, **child users** are provided with a means of actively confirming whether to receive a **direct message** from a **user** before it is visible to them, unless **direct messaging** is a necessary and time critical element of another functionality, in which case **child users** should be presented with a means of actively confirming before any interaction associated with that functionality begins.

A7.4 If the service has **automated location information displays**, the provider should implement default settings ensuring that **automated location information displays** are switched off for child users.

7B. Support for child users

Application

A7.5 This measure applies to a **provider** in respect of each **service** it provides that is either of the following, to the extent that it has an existing means of identifying **child users** of the service concerned:

a) at high **risk** of grooming; or
b) a **large service** that is at medium **risk** of grooming,

Recommendation

A7.6 Before a child user disables a **default setting** set out in Recommendation 7A, the **provider** should provide information to **child users** regarding the potential risk involved. The information should assist **child users** in understanding the implications of disabling that default setting, including the protections it affords.

A7.7 The **provider** should provide the following information when a **child user** seeks to respond to a request sent by another **user** via a **user connection**, before the connection is finalised:

a) the types of interactions that would be enabled through establishing a connection; and
b) the options available to take action against a **user**, such as **blocking**, **muting**, reporting or equivalent action.

A7.8 The **provider** should provide the following information when a **child user** exchanges a **direct message** with a **user** for the first time:

a) a reminder that this is the first direct communication with that user; and
b) the options available to take action against a **user**, such as **blocking**, **muting**, reporting or equivalent action,

A7.9 unless **direct messaging** is a necessary and time critical element of another functionality, in which case the **child user** may be provided this information before any interaction associated with that functionality begins.

A7.10 The **provider** should provide the following information when a **child user** seeks to block, mute, report, or take equivalent action against a **user**:

a) the effect of the action, including the types of interactions that it would restrict and whether the **user** would be notified; and
b) the further options available to limit interaction with the **user** or increase their safety.

A7.11 The **provider** should ensure that the information provided in line with paragraphs A7.6 to A7.9 is:

a) prominently displayed to **child users;** and
b) clear, comprehensible and easy for a **child user** to understand.

A8. Recommender system testing

8A. Safety metrics for on-platform testing of recommender systems

Application

A8.1 This measure applies to a **provider** in respect of each **service** it provides that meets both of the following conditions:

a) the provider conducts on-platform testing of **recommender systems** on the service; and
b) the **service** is at medium or high **risk** of at least two of the following **kinds of illegal harm**:

i) terrorism;
ii) CSAM;
iii) encouraging or assisting suicide (or attempted suicide) or serious self-harm;
iv) hate;
v) harassment, stalking, threats and abuse;
vi) drugs and psychoactive substances;
vii) extreme pornography;
viii) intimate image abuse; or
ix) the foreign interference offence.

Recommendation

A8.2 The **provider** should produce and analyse safety metrics when conducting **on-platform testing** of an actual or proposed **recommender system design change.**

A8.3 The safety metrics should enable the **provider** to understand whether a **recommender system design change** would increase the risk of **users encountering illegal content**, compared with the existing variant of the **recommender system** and include the following (or equivalent):

a) the total number of individual items of **content** that are assessed and identified as **illegal content** or as an **illegal content proxy** in response to a complaint made during the testing period; and
b) for each such item:

i) the number of times that **content** was displayed to **users** (impressions);
ii) the number of unique **users** that the **content** was displayed to (reach).

A8.4 The **provider** should ensure that:

a) the testing environment is set up in a way that enables the processing of complaints about **content** suspected to be **illegal content** or an **illegal content proxy**;
b) the period during which **on-platform testing** is conducted is sufficient so as to allow for complaints to be received; and
c) information that is relevant to producing the safety metrics in paragraph A8.3 is retained during the testing period.

A8.5 The **provider** should maintain a log of the results of each on-platform test, which should include a record of:

a) the safety metrics produced against each variant of the **recommender system** tested;
b) a description of each variant of the **recommender system**, including its respective design characteristics; and
c) the design decision taken on which variant of the **recommender system** to deploy following **on-platform testing**.

A8.6 The **provider** should ensure that the log is:

a) made available and is easily accessible to staff involved directly or indirectly in the development and testing of **recommender systems**; and
b) referred to by relevant staff in the context of future **recommender system design changes**.

A9. Enhanced user controls

9A. User blocking and muting

Application

A9.1 This measure applies to a **provider** in respect of each **large service** it provides and in respect of which each of the following conditions are met:

a) the **service** is at medium or high **risk** of one or more of the following **kinds of illegal harm**:

i) grooming;
ii) encouraging or assisting suicide (or attempted suicide) or serious self-harm;
iii) hate;
iv) harassment, stalking, threats and abuse;
v) controlling or coercive behaviour;

b) the **service** has **user profiles**; and
c) the **service** has at least one of the following functionalities:

i) **user connection**;
ii) **posting content**;
iii) **user communication**, including but not limited to: (1) **direct messaging**; and (2) **commenting on content**

Recommendation

A9.2 The **provider** should offer every registered **user** options to **block** other **user accounts** on the **service**. A **user** (the "**blocking user**") should have the option to **block** each of:

a) a specific **user account**, whether or not **connected** to the **blocking user**; and
b) all **user accounts** which are not connected to the **blocking user**

(referred to together as "**blocked users**").

A9.3 **Blocking** means that:

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 (regardless of where on the service it is posted) and vice versa, including but not limited to: (i) **reactions** to and ratings of **content posted** by the **blocked user**; and (ii) **content** originally posted by a **blocked user** which is subsequently posted by another **user**; and
c) the **blocking user** and **blocked user**, if they were **connected**, will no longer be **connected**.

A9.4 The **provider** should offer every registered **user** the option to **mute** other **user accounts** on the relevant **service**. A **user** (the "**muting user**") should have the option to **mute** individual **user accounts** ("**muted users**"), whether or not **connected** to the **muting user**.

A9.5 **Muting** means that the **muting user** will not **encounter** any **content posted** by **muted users** on the **service**, including: (1) **reactions** to and ratings of **content posted** by the **muted user**; and (2) **content** originally posted by a **muted user** which is posted by another **user**, unless

the **muting user** visits the **user profile** of a **muted user**, in which case the **muting user** will experience that **muted user's user profile** as if they had not **muted** the **muted user**.

A9.6 For the avoidance of doubt:

a) **muted users** should continue to **encounter** the **muting user's content**;
b) functionality from the **muted users**' perspective should continue as if the **muting user** has not **muted** them; and
c) **providers** should not at any time notify **muted users** or otherwise make them aware that they have been **muted** by the **muting user**.

9B. Disabling comments

Application

A9.7 This measure applies to a **provider** in respect of each **large service** it provides and in respect of which the following conditions are met:

a) the **service** is at high or medium **risk** of one or more of the following kinds of illegal harms

i) grooming;
ii) encouraging or assisting suicide (or attempted suicide) or serious self-harm;
iii) hate;
iv) harassment, stalking, threats and abuse; and

b) the **service** has the functionality of **commenting on content**.

Recommendation

A9.8 The **provider** should offer every registered **user** the option of preventing any other **users** of the **service** from **commenting on content** posted on the **service** by the **user** concerned.

A9.9 Registered **users** should be able to exercise the option referred to above:

a) when **posting content**; and
b) after having **posted content**.

9C. User verification/labelling schemes

Application

A9.10 This measure applies to a **provider** in respect of each **large service** it provides and in respect of which the following conditions are met:

a) the **service** is at high or medium **risk** of either or both of: (i) fraud **;** or (ii) the foreign interference offence; and
b) the **service** labels **user profiles** under one or more of the following: (i) a **notable user scheme**; or (ii) a **monetised scheme**.

A9.11 In this measure: (a) **notable user schemes** and **monetised schemes** are referred to together as "**relevant schemes**"; and (b) a **user** whose **user profile** is labelled under a **relevant scheme** is referred to as a "**relevant user**".

Recommendation

A9.12 The **provider** should have, and consistently apply, internal documented policies regarding the operation of **relevant schemes** on the service, which should, at a minimum:

a) be designed to reduce any **risk** of harm to users from fraud and/or the foreign interference offence associated with a **relevant scheme**, as identified in the **risk assessment** of the **service**;

b) set out: (1) the process for considering; and (2) the criteria and thresholds for deciding whether to:

i) label a **user profile**; and

ii) remove the label from the **user profile** of a **relevant user**.

In respect of a **notable user scheme**, the criteria and thresholds should set out how the **provider** will satisfy itself that:

iii) the **user account** of a **relevant user** is operated by or on behalf of the person by whom or on whose behalf it is held out as being operated; and

iv) if that person is held out as holding a particular position or role, that they hold that position or role;

c) set out safeguards to ensure that the **user profile** information (such as username and 'bio' text) provided by the **relevant user** when their **user profile** was labelled under a **notable user scheme** is not modified without the **provider** reviewing and consenting to that change;

d) set out the frequency with which and the circumstances in which the **provider** will conduct reviews to confirm whether the **user profiles** of **relevant users** continue to qualify to be labelled;

e) set out how the **provider** will treat **relevant users** and the **content** they post on the **service**, including **recommender systems**, content curation, user reporting and complaints, quality assurance, fact checking, content moderation, and account security;

f) be communicated to relevant staff, including through regular training (in particular when a policy is modified); and

g) be regularly reviewed and updated to ensure the policy remains fit for purpose. As part of regularly reviewing the policy, the **provider** should, if it considers it appropriate, take into account one or more of the following: user feedback and reporting; user experience testing; and engagement with persons with relevant expertise.

A9.13 The **provider** should provide:

a) the following information on the **user profile** of a **relevant user**:

i) why the **user profile** is labelled; and

ii) if the **provider** operates more than one type of **relevant scheme** on the **service**, the **relevant scheme(s)** under which the **user profile** is labelled; and

b) a user-facing description of the **relevant scheme(s)**, which should:

i) be in writing;

ii) be clear and accessible, for the purposes of which providers should act in accordance with the factors set out in Recommendation 6B (terms of service);

iii) explain how and why user profiles are labelled (including different categories of labelling and, in particular, specifying whether a relevant scheme is or is not a notable user scheme);

iv) explain how and why relevant users may have a label removed from their user profile; and

v) be consistent with (but need not include) every detail of internal policies.

A10. User access

10A. Removing accounts of proscribed organisations

Application

A10.1 This measure applies to a **provider** in respect of each **service** it provides.

Recommendation

A10.2 Where a **provider** has:

a) reviewed content which is **proscribed organisation content** or amounts to a breach of an equivalent standard as set out in the service's **terms of service** (referred to in this measure as "relevant content"); or
b) received a user complaint or has otherwise been made aware that a **user account** on a service may be operated by or on behalf of a **proscribed organisation**,

it should consider whether the **user account** that (1) posted the relevant content or (2) is the subject of the user complaint might be operated by or on behalf of a **proscribed organisation**.

A10.3 The provider should remove a **user account** from a **service** when they have reasonable grounds to infer it is operated by or on behalf of a **proscribed organisation**.

A10.4 Reasonable grounds to infer a **user account** is operated by or on behalf of a **proscribed organisation** may arise where at least two of the following are true of the **user profile**:

a) the username is the same as that of: (1) a **proscribed organisation**; or (2) an alias as specified in an order of the Secretary of State;[5]
b) the **user profile** image or any end-user configurable image setting is **proscribed organisation content**;
c) the **user profile** information, such as 'bio' text or other descriptive text, is **proscribed organisation content**.

A10.5 Reasonable grounds may also arise where one or none of the above is true, but where a significant proportion of a reasonably sized sample of the **content** recently posted by the **user account** is **proscribed organisation content**. What amounts to a reasonable sample size will depend on the amount of **content posted** by the account and the nature of the **service**. "Content recently posted by the account" refers to the newest **content posted** by the **user account** irrespective of date, rather than the **content posted** by the **user account** in a recent date range.

A10.6 For the purposes of the preceding paragraph, "content" does not include **content** that has been privately communicated,[6] unless the **provider** has explicit consent to view the **content** in question.

[5] Under section 3(6) of the Terrorism Act 2000.
[6] Ofcom has published [*draft*] **guidance on content communicated "publicly" and "privately" under the Online Safety Act**. (This footnote is for information and does not form part of the Code of Practice.)

Safeguards for freedom of expression and privacy

A10.7 Recommendations 5E(i) and (ii) (appeals) are safeguards to protect users' right to freedom of expression.

A10.8 Paragraph A10.6 is a safeguard to protect users' right to privacy.

A11. Definitions and interpretation

Definitions

A11.1 In these Codes, terms in bold have the meanings given below.[7]

A11.2 Definitions which are taken from the **Act** are reproduced here for ease of reading, but marked in *italics* together with a reference to the relevant section number. In case of any conflict, between the **Act** and italicised definition, the definition in the **Act** takes precedence. Omission of any term defined in the **Act** from the list below is not intended to imply that it does not apply.

A11.3 Where non-italicised terms are defined differently below than in the **Act**, the definition below applies

A11.4

Term	Meaning
Act	The Online Safety Act 2023
Affected person	***Section 20(5)*** *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.*

[7] In some cases, the terms defined in these Codes are also used in the Register of Risks. For the avoidance of doubt, the definitions set out in this section apply to the measures in these Codes.

Term	Meaning
Appeal	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**.
Automated location information display	A function which automatically creates and displays to other **users** the location information of **users**, including via the following (where relevant): a) shared **content**; b) **user profile**; and c) functionalities that display the live location information of users. For the purposes of this definition, location information about a **user** may be generated automatically by a **provider** using GPS data, local wi-fi or other means.
Block	As defined in paragraph A9.3 of Recommendation 9A (user blocking and muting).
Blocked user	For the purposes of Recommendation 9A (user blocking and muting), as defined in paragraph A9.2.

Term	Meaning
Blocking user	For the purposes of Recommendation 9A (user blocking and muting), as defined in paragraph A9.2.
Child user	A **user** who is under the age of 18.
Combined service	***Section 4(7)*** *A regulated user-to-user service that includes a public search engine.*
Commenting on content	**User-to-user service** functionality that allows users to reply to **content**, or post **content** in response to another piece of **content**, visually accessible directly from the original **content** without navigating away from that **content**.
Connected	Describes where two users have a **user connection**.
Connection lists	A list of a **user**'s connections which is visible to other **users** via a **user profile**. This may include 'friends', 'followers', 'subscribers' or indications of mutual connections.
Content	***Section 236(1)*** *Anything that is communicated by means of an internet service, whether publicly or privately, including written material or messages, oral communications, photographs, videos, visual images, music and data of any description.* For the avoidance of doubt, comments, titles and descriptions are considered to be 'content' within this definition, as are livestreaming videos or audio, and hyperlinks.

Term	Meaning
Content moderation function	Systems or processes designed to swiftly take down **illegal content** of which a service is aware.
CSAM (child sexual abuse material)	Content that amounts to an offence specified in any of the following paragraphs of Schedule 6 to the **Act**— a) paragraph 1 to 4, 7 or 8, or paragraph 10 so far as any of the inchoate offences relate to an offence specified in those paragraphs; or b) paragraph 9, or paragraph 13 so far as any of the inchoate offences relate to an offence specified in that paragraph.
CSEA content	***Section 59(9)*** *Content that amounts to an offence specified in Schedule 6 [to the Act].*
Default settings	Automatic settings for functionalities set by a **service** that can be disabled by a **user**.
Detected content	**Content** detected by the use of a **relevant technology** as being (or as likely to be) **target content** (and related expressions are to be read accordingly).
Direct messaging	**User-to-user service** functionality that allows a **user** to send a message to one recipient at a time and which can only be immediately viewed or read by that specific recipient.
Encounter (in relation to content)	***Section 236(1)*** *Read, view, hear or otherwise experience content.*

Term	Meaning
False positive	**Detected content** that is not **target content**.
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.
Fuzzy keyword detection technology	Technology which analyses **content** to assess whether it contains one or more **keywords**, or text which is similar to one or more **keywords** (for example, because that text is an abbreviation or common misspelling of those **keywords**, and/or includes some replacement characters). This does not include technology which identifies text that is similar to one or more keywords through the use of machine learning.
Governance body	A body which makes decisions within an organisation. These may vary by organisation type and size, but boards of directors are commonly the most senior governance forums in corporations.

Term	Meaning
Illegal content	***Section 59*** *Content that amounts to a relevant offence.* *Content consisting of certain words, images, speech or sounds amounts to a relevant offence if—* *a) the use of the words, images, speech or sounds amounts to a relevant offence,* *b) the possession, viewing or accessing of the content constitutes a relevant offence, or* *c) the publication or dissemination of the content constitutes a relevant offence.* (For guidance on determining when **content** 'amounts to' a relevant offence, refer to the Illegal Content Judgements Guidance.)
Illegal content proxy	**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**).
Illegal content safety duties	The duties in section 10 of the **Act**.
Illegal harm	Harms arising from **illegal content** and the commission and facilitation of **priority offences**.

Term	Meaning
Internet service	***Section 228(1)*** *A service that is made available by means of the internet.* [*See also the rest of the section.*]
Keyword	A string of text (such as a word, phrase, special character, hashtag or abbreviation) included in a keyword list.
Kind of illegal harm	Shall be interpreted in accordance with the section headed 'Risk and risk of harm' below (which begins at paragraph A11.4).
Large service	A **service** which has more than 7 million **monthly United Kingdom users** (See also paragraphs A11.6 to A11.10)
Monetised scheme	A scheme by which a **service** labels the **user profile** 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 **monetised scheme** may appear on that user's profile and/or any content they publish. Services may or may not refer to such schemes as "verification" schemes.
Monthly (in relation to a number of United Kingdom users)	See paragraphs A11.6 to A11.10.

Term	Meaning
Multi-risk service	A **service** assessed as being at least medium risk in relation to at least two kinds of **priority offences** in accordance with the section headed 'Risk and risk of harm' below (which begins at paragraph A11.4). For the purposes of this definition, rows 2, 2A and 2B are to be treated as one kind of priority offence.
Mute	As defined in paragraph A9.5 of Recommendation 9A (user blocking and muting).
Muted user	For the purposes of Recommendation 9A (user blocking and muting), as defined in paragraph A9.4.
Muting user	For the purposes of Recommendation 9A (user blocking and muting), as defined in paragraph A9.4.
Network expansion prompts	A functionality that, by means of a network recommender system, recommends **users** or groups for other **users** to connect with. This can include specific **users** who have similar interests, who are close geographically, who attend the same school or workplace, or with whom a user has a mutual connection.

Term	Meaning
Non-connected user	For the purposes of Recommendation 7A (safety defaults for child users), a user with which another user has not established a valid connection. In the case in which the connection is between a **child user** and a non-**child user**, a valid connection requires the **child user** to initiate a connection or confirm a connection via 'friending', 'following', or 'subscribing' to the non-**child user**. In the case in which the connection is between two **child users** a valid connection requires both **child users** to confirm a connection via 'friending' or reciprocated 'following' or 'subscribing'.
Notable user 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.

Term	Meaning
Offence concerning articles for use in frauds	An offence specified in any of the following paragraphs of Schedule 7 to the **Act**: paragraph 29(c) and paragraph 30, or paragraph 33 so far as any of the inchoate offences relate to an offence specified in paragraphs 29(c) or 30.
On-platform testing	The process of live testing the operation of different variants of a **recommender system** on a service across control group and treatment groups comprised of **users** of the **service**. It involves the collection of data to produce metrics relating to certain identified factors, such as commercial or user safety. The methods may include (but are not limited to): • A/B/x testing: a randomised control trial in which a treatment group is served content from an update version of a recommender system, and a control group is served content from the current recommender system, with a view to comparing their performance against identified metrics. • Multi Arm Bandit Testing: a randomised control trial employing machine learning techniques to allocate users to the "best" performing version of a recommender system during the course of the testing period.

Term	Meaning
Perceptual hash matching technology	Image matching technology which compares the similarity between hashes created from images by means of an algorithm known as a perceptual hash function, to assess whether those images are perceptually similar to each other. This does not include technology which compares similarity through the use of machine learning.
Posting content	**User-to-user service** functionality allowing users to upload and share **content** on open channels of communication.
Precision	A measure of statistical accuracy, calculated as the proportion of **detected content** that a **relevant technology** has correctly identified as **target content**.
Priority illegal content	***Section 59(10)*** *(a) terrorism content, (b) CSEA content, and (c) content that amounts to an offence specified in Schedule 7.*
Priority offences	The offences set out in Schedules 5, 6 and 7 to the Act.

Proactive technology	***Section 231*** *Means—* *a) content identification technology* *b) user profiling technology, or* *c) behaviour identification technology,* *but this is subject to subsections (3) and (7).* *(2) "Content identification technology" means 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).* *(3) But content identification technology is not to be regarded as proactive technology if it is used in response to a report from a user or other person about particular content.* *(4) "User profiling technology" means technology which analyses (any or all of)—* *a) relevant content* *b) user data, or* *c) metadata relating to relevant content or user data,* *for the purposes of building a profile of a user to assess characteristics such as age.* *(5) Technology which—* *a) analyses data specifically provided by a user for the purposes of the provider assessing or establishing the user's age in order to decide whether to allow the user to access*

Term	Meaning
	a service (or part of a service) or particular content, and *b) does not analyse any other data or content, is not to be regarded as user profiling technology.* *(6) "Behaviour identification technology" means technology which analyses (any or all of)—* *a) relevant content,* *b) user data, or* *c) 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).* *(7) But behaviour identification technology is not to be regarded as proactive technology if it is used in response to concerns identified by another person or an automated tool about a particular user.* *[See also the rest of the section.]*
Proscribed organisation	A group or organisation proscribed by the Secretary of State under section 3 of the Terrorism Act 2000.[8]

[8] As contained in Schedule 2 to the Terrorism Act 2000 and set out in the Government's list of proscribed terrorist groups or organisations.

Term	Meaning
Proscribed organisation content	**Content** which amounts to an offence specified in any of the following paragraphs of Schedule 5 to the **Act**: a) paragraphs 1(a) to (e); b) paragraphs 1(f) to (p) and 3, where the "terrorism" for the purpose of the offence is an action taken for the benefit of a **proscribed organisation**; or c) paragraph 4 so far as any of the inchoate offences relate to an offence falling within points (a) or (b) above.
Provider	A provider of a **regulated user-to-user service**.
Reacting to content	A **user-to-user service** functionality. Described by user communication functionality type. Includes functionalities such as 'liking' or 'loving' **content**.
Recall	A measure of statistical accuracy, calculated as the proportion of **target content** analysed by a **relevant technology** that the technology has **detected**.

Term	Meaning
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, including but not limited to: a) **User** feedback, such as interactions with a piece of **content** by means of likes, views and shares; b) Predicted engagement with **content** based on their consumption history, such as likelihood of liking, sharing, and commenting on a piece of **content**; c) Profile and contextual characteristics, such as age and location; d) **Content** liked by **users** with a similar consumption and engagement history; e) Popularity of a certain piece of **content**. References to recommender systems in these Codes do not include those employed by **providers** in search functionalities or network recommender systems that suggest **users** and groups to follow.

Term	Meaning
Recommender system design change	Any small and incremental alterations to the design of an existing **recommender system** that are made as part of product management, such as changes to the identified pool of **content**, types and weighting of factors used by the machine learning model(s). It does not include design changes that: • would amount to a significant change and therefore trigger a risk assessment under section 8(4) of the Act; or • are made in connection with a live response to a national security threat or other emergency; • would not be deployed for UK users of the service.

Term	Meaning
Regulated user-generated content	***Section 55*** *In relation to a regulated user-to-user service, means user-generated content, except—* *a) emails,* *b) SMS messages,* *c) MMS messages,* *d) one-to-one live aural communications (see subsection (5) [of section 55]),* *e) comments and reviews on provider content (see subsection (6) [of section 55]),* *f) identifying content that accompanies content within any of paragraphs (a) to (e), and* g) *news publisher content (see subsection (8) [of section 55]).*
Regulated user-to-user service	A **user-to-user service** as defined in section 3 of the **Act**, which is a regulated user-to-user service under section 4 of the **Act** (subject to the disapplication in section 5 of the **Act**).

Relevant complaints	The following kinds of complaint: a) complaints (including **reports**) 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** if they consider that the provider is not complying with its duties in relation to **illegal content**, content reporting, freedom of expression or privacy); 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 **user-to-user 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

Term	Meaning
	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).
Relevant scheme(s)	For the purposes of Recommendation 9C (user verification/labelling schemes), as defined in paragraph A9.11.
Relevant technology	The kind of technology specified in the measure in question.
Relevant user	For the purposes of Recommendation 9C (user verification/labelling schemes), as defined in paragraph A9.11.
Reporting and complaints duties	The duties in sections 20 and 21 of the **Act**.
Reports	Complaints by **users** and **affected persons** about content present on a **service** which they consider to be **illegal content**, made using a reporting function or tool provided by the **service.**

Term	Meaning
Risk	Shall be interpreted in accordance with the section headed 'Risk and risk of harm' below (which begins at paragraph A11.4).
Risk assessment	The most recent risk assessment carried out by the provider pursuant to section 9 of the **Act.**
Service	A **regulated user-to-user service.**
Taking down	***Section 236(1)*** *any reference to taking down content is to any action that results in content being removed from a user-to-user service or being permanently hidden so users of the service cannot encounter it (and related expressions are to be read accordingly);*
Target content	**Content** of the kind the use of a **relevant technology** is designed to identify.
Terms of service	***Section 236(1)*** *all documents (whatever they are called) comprising the contract for use of the service (or of part of it) by United Kingdom users*
Terrorism content	***Section 59(8)*** *Content that amounts to an offence specified in Schedule 5 [to the Act].*
Trusted flagger	Save as defined in Recommendation 5I for the purposes of that recommendation, means an entity set out in Recommendation 5I for which the provider has established a dedicated reporting channel.

Term	Meaning
User-to-user service	***Section 3(1)*** *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.* [*See also* ***section 3(2)***]
United Kingdom user	***Section 227(1)*** *a) where the user is an individual, the individual is in the United Kingdom;* **b)** *where the user is an entity, the entity is incorporated or formed under the law of any part of the United Kingdom.*
URL	Uniform Resource Locator, meaning a reference that specifies the location of a resource accessible by means of the internet.

Term	Meaning
User	***Section 227*** *(1) [See definition of **United Kingdom user** above]* *(2) For the purposes of references in this Act to a user of a service it does not matter whether a person is registered to use a service.* *(3) References in this Act to a user of a service do not include references to any of the following when acting in the course of the provider's business—* *a) where the provider of the service is an individual or individuals, that individual or those individuals;* *b) where the provider is an entity, officers of the entity;* *c) persons who work for the provider (including as employees or volunteers);* *d) any other person providing a business service to the provider such as a contractor, consultant or auditor.* *(4) [defines "acting in the course of the provider's business"]* *(5) [defines "service"]* *(6) [defines "officer"]*

Term	Meaning
User account	Representations of a **user** in a service's information system. They may contain information required for registration to a particular **service** that are often attributes of a **user's** identity such as name, age, contact details and preferences.
User communication	**User-to-user service** functionality type that describes functionalities by means of which **users** can communicate with one another either synchronously or asynchronously. Includes communication across open and closed channels.
User connection	A functionality that allows **users** to follow or subscribe to other **users**. **Users** must sometimes be connected to view all or some of the **content** that each **user** shares.
User profiles	**User-to-user service** functionality that represents a collection of identifying information about a **user** conveyed to other **users** of the **service**. This can include information that may be displayed to other **users** such as images, usernames, and biographies.
User-to-user part (of a service)	***Section 236(1)*** *… the part of the [user-to-user] service on which content that is user-generated content in relation to the service is present.*

Term	Meaning
Volunteer	A person involved in content moderation who, in relation to that involvement, is not: a) employed by the **provider** or anyone else, b) remunerated, or ***c)*** acting by way of a business.

Risk and risk of harm

A11.5 A service is at medium or high risk of a kind of illegal harm specified in the table if the **risk assessment** of the service identified a medium or high risk (as the case may be) in relation to the offences (taken together) specified in the table in relation to that harm, including (where relevant) as further specified in the table.

A11.6 In relation to each priority offence listed in rows 2 to 16, the offence also includes the priority offences of encouraging, assisting, conspiring to commit, aiding, abetting, counselling, procuring, attempting, or, (in Scotland), inciting or being involved art and part in the commission of that offence. The offences are priority offences unless otherwise specified.

	Kind of illegal harm	Offences
1.	Terrorism	An offence specified in Schedule 5 to the Act.
2.	CSAM	An offence specified in any of paragraphs 1 to 4, 7, 8 or 10 of Schedule 6 to the **Act**.
2A.	Image-based CSAM	An offence specified in any of paragraphs 1 to 4, 7, 8 or 10 of Schedule 6 to the **Act**, so far as the risk in relation to those offences relates to **CSAM** in the form of photographs, videos or visual images.

	Kind of illegal harm	Offences
2B.	CSAM URLs	An offence specified in any of paragraphs 1 to 4, 7, 8 or 10 of Schedule 6 to the **Act**, so far as the risk in relation to those offences relates to users encountering **CSAM** by means of or facilitated by **URLs** present on the service.
3.	Grooming	An offence specified in any of paragraphs 5, 6, 11 or 12 of Schedule 6 to the **Act**.
4.	Encouraging or assisting suicide (or attempted suicide) or serious self-harm	An offence under: (a) section 2 of the Suicide Act 1961 (assisting suicide etc); (b) section 13 of the Criminal Justice Act (Northern Ireland) 1966 (c. 20 (N.I.)) (assisting suicide etc); (c) section 184 of the Online Safety Act 2023 (a relevant non-priority offence).

	Kind of illegal harm	Offences
5.	Hate	An offence under any of the following provisions of the Public Order Act 1986— (a) section 18 (use of words or behaviour or display of written material); (b) section 19 (publishing or distributing written material); (c) section 21 (distributing, showing or playing a recording); (d) section 29B (use of words or behaviour or display of written material); (e) section 29C (publishing or distributing written material); (f) section 29E (distributing, showing or playing a recording). An offence under any of the following provisions of the Crime and Disorder Act 1998— (a) section 31 (racially or religiously aggravated public order offences); (b) section 32 (racially or religiously aggravated harassment etc). An offence under section 50A of the Criminal Law (Consolidation) (Scotland) Act 1995 (racially-aggravated harassment).

	Kind of illegal harm	Offences
6.	Harassment, stalking, threats and abuse	An offence under section 16 of the Offences against the Person Act 1861 (threats to kill). An offence under any of the following provisions of the Public Order Act 1986— (a) section 4 (fear or provocation of violence); (b) section 4A (intentional harassment, alarm or distress); (c) section 5 (harassment, alarm or distress). An offence under any of the following provisions of the Protection from Harassment Act 1997— (a) section 2 (harassment); (b) section 2A (stalking); (c) section 4 (putting people in fear of violence); (d) section 4A (stalking involving fear of violence or serious alarm or distress). An offence under any of the following provisions of the Protection from Harassment (Northern Ireland) Order 1997 (S.I. 1997/1180 (N.I. 9))— (a) Article 4 (harassment); (b) Article 6 (putting people in fear of violence) An offence under any of the following provisions of the Criminal Justice and Licensing (Scotland) Act 2010 (asp 13)— (a) section 38 (threatening or abusive behaviour); (b) section 39 (stalking).
7.	Controlling or coercive behaviour	An offence under section 76 of the Serious Crime Act 2015 (controlling or coercive behaviour in an intimate or family relationship).

	Kind of illegal harm	Offences
8.	Drugs and psychoactive substances	An offence under any of the following provisions of the Misuse of Drugs Act 1971— (a) section 4(3) (unlawful supply, or offer to supply, of controlled drugs); (b) section 9A (prohibition of supply etc of articles for administering or preparing controlled drugs); (c) section 19 (inciting any other offence under that Act). An offence under section 5 of the Psychoactive Substances Act 2016 (supplying, or offering to supply, a psychoactive substance).

9.	Firearms and other weapons	An offence under section 1(1) or (2) of the Restriction of Offensive Weapons Act 1959 (sale etc of flick knife etc). An offence under any of the following provisions of the Firearms Act 1968— (a) section 1(1) (purchase etc of firearms or ammunition without certificate); (b) section 2(1) (purchase etc of shot gun without certificate); (c) section 3(1) (dealing etc in firearms or ammunition by way of trade or business without being registered); (d) section 3(2) (sale etc of firearms or ammunition to person other than registered dealer); (e) section 5(1), (1A) or (2A) (purchase, sale etc of prohibited weapons); (f) section 21(5) (sale etc of firearms or ammunition to persons previously convicted of crime); (g) section 22(1) (purchase etc of firearms or ammunition by person under 18); (h) section 24 (supplying firearms to minors); (i) section 24A (supplying imitation firearms to minors). An offence under any of the following provisions of the Crossbows Act 1987— (a) section 1 (sale and letting on hire of crossbow); (b) section 2 (purchase and hiring of crossbow). An offence under any of the following provisions of the Criminal Justice Act 1988— (a) section 141(1) or (4) (sale etc of offensive weapons); (b) section 141A (sale of knives etc to persons under 18).

	Kind of illegal harm	Offences
		An offence under any of the following provisions of the Criminal Justice (Northern Ireland) Order 1996 (S.I. 1996/3160 (N.I. 24))— (a) Article 53 (sale etc of knives); (b) Article 54 (sale of knives etc to minors). An offence under any of the following provisions of the Knives Act 1997— (a) section 1 (unlawful marketing of knives); (b) section 2 (publication of material in connection with marketing of knives). An offence under any of the following provisions of the Firearms (Northern Ireland) Order 2004 (S.I. 2004/702 (N.I. 3))— (a) Article 24 (sale etc of firearms or ammunition without certificate); (b) Article 37(1) (sale etc of firearms or ammunition to person without certificate etc); (c) Article 45(1) or (2) (purchase, sale etc of prohibited weapons); (d) Article 63(8) (sale etc of firearms or ammunition to people who have been in prison etc); (e) Article 66A (supplying imitation firearms to minors). An offence under section 36(1)(c) or (d) of the Violent Crime Reduction Act 2006 (sale etc of realistic imitation firearms). An offence under any of the following provisions of the Air Weapons and Licensing (Scotland) Act 2015 (asp 10)— (a) section 2 (requirement for air weapon certificate); (b) section 24 (restrictions on sale etc of air weapons).

	Kind of illegal harm	Offences
10.	Unlawful immigration and human trafficking	An offence under any of the following provisions of the Immigration Act 1971— (a) section 24(A1), (B1), (C1) or (D1) (illegal entry and similar offences); (b) section 25 (assisting unlawful immigration). An offence under section 2 of the Modern Slavery Act 2015 (human trafficking). An offence under section 1 of the Human Trafficking and Exploitation (Scotland) Act 2015 (asp 12) (human trafficking). An offence under section 2 of the Human Trafficking and Exploitation (Criminal Justice and Support for Victims) Act (Northern Ireland) 2015 (c. 2 (N.I.)) (human trafficking).
11.	Sexual exploitation of adults	An offence under any of the following provisions of the Sexual Offences Act 2003— (a) section 52 (causing or inciting prostitution for gain); (b) section 53 (controlling prostitution for gain). An offence under any of the following provisions of the Sexual Offences (Northern Ireland) Order 2008 (S.I. 2008/1769 (N.I. 2))— (a) Article 62 (causing or inciting prostitution for gain); (b) Article 63 (controlling prostitution for gain).
12.	Extreme pornography	An offence under section 63 of the Criminal Justice and Immigration Act 2008 (possession of extreme pornographic images).

	Kind of illegal harm	Offences
13.	Intimate image abuse	An offence under section 33 of the Criminal Justice and Courts Act 2015 (disclosing, or threatening to disclose, private sexual photographs and films with intent to cause distress) [*OR, if section 188 of the Online Safety Act is brought into force and Schedule 7 to the Act is amended accordingly before we issue our final document, section 66B of the Sexual Offences Act 2003]*. An offence under section 2 of the Abusive Behaviour and Sexual Harm (Scotland) Act 2016 (asp 22) (disclosing, or threatening to disclose, an intimate photograph or film).
14.	Proceeds of crime	An offence under any of the following provisions of the Proceeds of Crime Act 2002— (a) section 327 (concealing etc criminal property); (b) section 328 (arrangements facilitating acquisition etc of criminal property); (c) section 329 (acquisition, use and possession of criminal property).

	Kind of illegal harm	Offences
15.	Fraud (and financial services)	An offence under any of the following provisions of the Fraud Act 2006— (a) section 2 (fraud by false representation); (b) section 4 (fraud by abuse of position); (c) section 7 (making or supplying articles for use in frauds); (d) section 9 (participating in fraudulent business carried on by sole trader etc). An offence under section 49(3) of the Criminal Justice and Licensing (Scotland) Act 2010 (articles for use in fraud). An offence under any of the following provisions of the Financial Services and Markets Act 2000— (a) section 23 (contravention of prohibition on carrying on regulated activity unless authorised or exempt); (b) section 24 (false claims to be authorised or exempt); (c) section 25 (contravention of restrictions on financial promotion). An offence under any of the following provisions of the Financial Services Act 2012— (a) section 89 (misleading statements); (b) section 90 (misleading impressions).
16.	Foreign interference offence	An offence under section 13 of the National Security Act 2023 (foreign interference).

User numbers

A11.7 This section applies for the purpose of determining whether a **service** is to be treated as having more than a particular number of monthly **United Kingdom users.**

A11.8 A **service** is to be so treated from such time as the number of monthly **United Kingdom users** of the **user-to-user part** of the service is more than the number in question.

A11.9 The **service** is to continue to be so treated until such time as the number of monthly **United Kingdom users** of the **user-to-user part** of the **service** is at or below the specified number for a continuous period of six months.

A11.10 Paragraph A11.7 may apply again to a **service** that has ceased to be so treated in accordance with paragraph A11.8.

A11.11 The number of monthly **United Kingdom users** of the **user-to-user part** of the **service** is the mean number of **United Kingdom users** per month, calculated for:

a) the period of 12 months ending with the month preceding the time in question; or
b) if the **service** not been in operation for that period, the period for which the service has operated.