Additional Safety Measures Statement

Statement

Published 16 November 2026

For more information on this publication, please visit ofcom.org.uk

Executive summary 3

1. Introduction, our duties and navigating the Statement 4
2. Livestreaming 13
3. User Sanctions 36
4. User controls 100
5. Highly effective age assurance in the Illegal Content Codes 113
6. Increasing the effectiveness of safety defaults and supportive messaging 131
7. Broadening content appeals 143
8. Introduction to Proactive Technology 155
9. Deployment of proactive technology to detect image-based CSAM 163
10. Hash matching 221
11. Recommender systems 258
12. Amendments to the Illegal Content Judgements Guidance for child sexual abuse material 283
13. Combined Impact Assessment 307
14. Statutory Tests 319

Executive summary

In June 2025, we consulted on a range of additions to the Codes of Practice, which set out the measures that service providers should take to protect people from harm online.[1]

We have already accelerated the implementation of two core measures.2 This statement sets out our decisions on the remaining measures covered by the June 2025 consultation.

The main theme of the decisions we are making today is protecting children from harm online. In particular, we are recommending that service providers should:[3]

- Use proactive technology capable of detecting image-based child sexual abuse material (CSAM) that hasn't previously been identified;
- Deploy a technology called hash matching to detect CSAM. We already recommend the use of hash matching in our Codes. Today we are expanding the range of services this recommendation applies to and are making clear that regulated user-to-user pornographic services should use this tool;
- Ban users who generate, share, or upload child sexual exploitation and abuse content (CSEA);
- Take action to remove illegal content from children's recommender feeds; and
- Strengthen the steps they are taking to protect children from grooming. The Codes already say that service providers should take steps to prevent strangers from contacting children. Today we are making clear that if they do not wish to apply these steps to all their users, they should use highly effective age assurance to work out which accounts should benefit from these protections.

As well as taking these steps to protect children, we are also introducing a number of additional measures designed to protect all users from illegal content. For example, we are recommending that service providers:

- Use hash matching technology to detect terrorism content; and
- Apply sanctions to UK users who share or upload illegal content.

Once these measures have passed through the necessary Parliamentary process, we will move quickly to drive compliance with them. These decisions are an important enhancement to our Codes and will materially improve online safety, in particular for children. However, more work is needed. Next Spring we will build on the decisions we have taken today and consult on proposals that service providers should use proactive technology capable of detecting a broader range of harms.

---

[1] The June 2025 Additional Safety Measures Consultation (June 2025 Consultation), this statement also includes decisions from the April 2025 User Controls Consultation (April 2025 Consultation).
[2] In May 2026, we introduced a requirement for certain services to use hash-matching technology to detect intimate image abuse content. In June 2026, we introduced a requirement for relevant services to develop protocols to respond to significant increases in illegal content during a crisis.
[3] Who each measure applies to is explained in the relevant chapters.

1. Introduction, our duties and navigating the Statement

What is this chapter about?

This chapter provides a high-level overview of the statement, including its background, our duties and functions, and the wider UK Government policy and legislative context. It also explains the scope of this statement, how it operates, and the next steps.

Background

1.1 The Online Safety Act 2023 (the Act) established Ofcom as the online safety regulator. The Act aims to make internet services safer for individuals in the UK by requiring regulated service providers to implement proportionate systems and processes to protect UK users from online harms.

1.2 As the online safety regulator, Ofcom must produce Codes of Practice setting out the actions that Ofcom recommends service providers should take to protect users. These Codes of Practice apply to a range of user-to-user and search services,[4] and are designed to protect all users from illegal content and child users from content which is harmful to them.[5] [6]

1.3 To begin implementing the Act, we published the December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement) and the April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement). These created a foundational basis for our online safety regime. Amongst other things, the December 2024 and April 2025 Statements issued Codes of Practice in relation to Illegal Content and Protection of Children (the Codes). Parliament subsequently approved these Codes, enabling them to come into force and providing Ofcom with the ability to enforce compliance with the relevant duties.

1.4 Building on these foundations, we consulted on proposals for a set of additional online safety measures across two publications. The first was our April 2025 User Controls Consultation (April 2025 Consultation). In this we proposed extending existing measures to allow users to block and mute accounts, and to disable comments to smaller service

[4] The Act defines a user-to-user service as an online service where content is generated on the service by a user, or uploaded to or shared on the service by a user, and this might be encountered by other users of the service. A search service refers to services that are, or include, a search engine that allows users to search more than one website or database.

[5] Illegal content is defined by section 59 of the Act as 'content that amounts to a relevant offence.' Relevant offences include both 'priority offences' and 'non-priority offences.' These are defined in the Act. Priority offences are the most serious offences covered by the Act, as defined by Parliament, and all providers need to prevent users from encountering content amounting to one of these offences. Ofcom (2024), Overview of Illegal Harms.

[6] Content harmful to children is content that presents a material risk of harm to a person under the age of 18. Section 53(1) of the Act requires us to give examples of content that we consider, or do not consider, to be content that is harmful to children. For more information see our April 2025 Guidance on Content Harmful to Children.

providers. The second was the June 2025 Additional Safety Measures Consultation (June 2025 Consultation). This included a wide range of measures to protect people in the UK from illegal content online and to protect children from content harmful to them. For more information on the aims and objectives of the Additional Safety Measures, see the next section titled 'Scope of this statement'.

1.5 We have already published our decisions relating to two proposals made in the June 2025 Consultation. In May 2026, we published our statement on Detecting Intimate Image Abuse (May 2026 Statement).[7] In June 2026, we published our statement on Crisis Response Protocols (June 2026 Statement).[8]

1.6 This document is a statement setting out our reasoning and decisions relating to the remaining proposals set out in the April 2025 and June 2025 Consultations.[9]

1.7 This chapter outlines our duties, functions and the scope of this statement, sets out next steps and explains how to navigate the document. We have included a Glossary of defined terms in annex 6: Glossary.

Scope of this statement

1.8 The decisions made in this statement will include a number of targeted safety measures in our Illegal Content and Protection of Children Codes.[10] The main theme of the decisions we are making today is protecting children from harm online. We are also introducing a number of additional measures designed to protect all users from illegal content.

1.9 This statement also sets out our updates to the Illegal Content Judgements Guidance, draft Guidance on Highly Effective Age Assurance for Part 3 services, draft Proactive Technology Guidance and draft Guidance on Appropriate Proportion of Human Review for hash matching. The Codes we are sending to the Secretary of State to be laid before Parliament include changes we explained in the June 2026 Statement on New priority offences. We will issue the draft guidance in final form when we issue the associated Codes of Practice.

Consultation process

1.10 The April 2025 and June 2025 Consultations included an open access response form for stakeholders to provide their views on our proposals. We received over 150 responses to both consultations from a broad range of stakeholders including: providers of regulated services, civil society organisations, public bodies, and individuals. We also received responses from three organised campaigns that were sent to us by over 530 individuals.

1.11 The June 2025 Consultation also included a wide programme of stakeholder engagement. Within the consultation period, we ran workshops and roundtables with stakeholders,

[7] The Codes for the May 2026 Statement have been issued and came into force on 9 September 2026.
[8] The Codes for the June 2026 Statement have been issued and came into force on 2 November 2026.
[9] We call our decision documents 'Statements' because they are statements of our reasoning for the decisions we have made.
[10] The measures set out in this statement will build on the Codes issued following our previous Online Safety Statements: December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement), April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement), May 2026 and June 2026 Statements.

including lived experience engagement workshops. These deliberative workshops were considered when arriving at our final decisions.

1.12 We thank stakeholders for their feedback across the consultations. We have addressed points that are specific to the decisions we have made on the measures in the relevant chapters and general stakeholder feedback is addressed in annex 1: Further stakeholder responses. We have considered all stakeholder responses received even where we have not referred to or quoted them in this statement.

How this statement works

1.13 This statement explains our decisions on the proposals set out in the April and June 2025 Consultations, including our consideration of stakeholder feedback and additional evidence.

1.14 The statement is organised by topic, with each chapter covering a related set of measures. The table of contents can be used to navigate directly to specific chapters and relevant measures.

1.15 Each chapter discusses the relevant decision and sets out:

a) Our decision
b) How the decision works
c) Our reasoning, including stakeholder feedback and how we reached our decision on the following points:
    i) benefits and effectiveness of the decision;
    ii) impact and costs of the decision;
    iii) rights assessment of the decision; and
    iv) who the decision applies to.

1.16 This reasoning, together with our Combined impact assessment in chapter 13 and our Equality and Welsh language impact assessment in annex 2, is our impact assessment.

How to navigate this statement

1.17 To ensure our statement is accessible and clear to navigate for a range of audiences, we have:

a) Published two summary documents:
    i) summary of our chapters which sets out a high-level outline of each chapter in our statement to help stakeholders navigate and engage with our document; and
    ii) summary of our decisions which includes tables outlining our decisions and which services they apply to.
b) Used a clear and signposted structure throughout;
c) Cross-referenced existing regulatory materials where possible; and
d) Included a summary at the start of each chapter setting out decisions and information for stakeholders.

Ofcom's duties and online safety functions

1.18 Ofcom is the independent regulator for communication services, with regulatory responsibilities for telecoms, post, broadcasting, and online safety. Our functions arise from a range of legislation. This includes the Communications Act 2003 (the 2003 Act) and the

Online Safety Act 2023 (the Act). We also have duties under the Equality Act 2021, the Human Rights Act 1998, and the Deregulation Act 2015.[11]

Communications Act 2003

1.19 The 2003 Act places duties on Ofcom that we must fulfil when exercising our regulatory functions, including our online safety functions. Our principal duty in carrying out our functions is:

a) to further the interests of citizens in relation to communication matters; and

b) to further the interests of consumers in relevant markets, where appropriate by promoting competition.[12]

1.20 By virtue of this, in relation to online safety, Ofcom is required to secure "adequate protection of citizens from harm presented by content on regulated services, through the appropriate use by providers of such services of systems and processes designed to reduce the risk of such harm".[13]

1.21 In performing our principal duty, we must have regard to principles set out in the 2003 Act, which says that regulatory activities should be transparent, accountable, proportionate, consistent and targeted only at cases where action is needed.[14]

1.22 The 2003 Act further requires that we have regard to the following factors (among other things) as they appear to us to be relevant in the circumstances.[15] [16] In making our decisions, we have considered factors including, but not limited to:

a) the risk of harm to UK citizens presented by regulated services;

b) the need for a higher level of protection for children than for adults;

c) the need for it to be clear to providers of regulated services how they may comply with their duties under the 2003 Act;

d) the need to exercise our functions to secure that providers may comply with such duties by taking or using measures, systems or processes which are proportionate to the size or capacity of the provider and the level of risk (and potential severity) of harm presented by the service;

e) the desirability of promoting the use of technologies which are designed to reduce the risk of harm to citizens; and

---

[11] For further information see chapter 14: Statutory tests (Schedule 4), chapter 13: Combined impact assessment, chapter 8: Introduction to proactive technology.
[12] Section 3(1) of the Communications Act 2003 (the 2003 Act).
[13] Section 3(2) of the 2003 Act.
[14] Section 3(3) of the 2003 Act. We must also have regard to any other principles appearing to us to represent best regulatory practice.
[15] Section 3(4A) of the 2003 Act.
[16] In relation to matters to which section 3(2)(g) is relevant. The 2003 Act sets out other matters to which Ofcom must, to the extent they appear to us relevant in the circumstances, have regard, in performing our duties. They include: the desirability of promoting competition and encouraging investment and innovation in relevant markets; the vulnerability of children and of others whose circumstances put them in need of special protection; the needs of persons with disabilities, the elderly and of those on low incomes; the desirability of preventing crime and disorder; the opinions of consumers and of members of the public generally; and the different interests of persons in the different parts of the UK and of the different ethnic communities within the UK. See chapter 14: Statutory tests.

f) the extent to which providers demonstrate, in a way that is transparent and accountable, that they are complying with their duties.

1.23 In line with our additional duties under the 2003 Act, we have also considered the vulnerability of children and of others whose circumstances put them in need of special protection. We have considered:

- the needs of disabled people, older people, and of those on low incomes;
- the opinions of consumers and of members of the public generally;
- the interests of persons in the different parts of the UK; and
- the interests of the different ethnic communities within the UK.

1.24 As a public authority, Ofcom must act lawfully, rationally, and fairly.

Online Safety Act 2023

1.25 The Act established Ofcom as the regulator for online safety and places a requirement on us to prepare and issue a set of Codes. These are a package of measures that Ofcom recommends for service providers to comply with their safety duties under the Act.

1.26 Under the Act, regulated service providers have safety duties.[17] They should identify, mitigate, and manage the risk of harm from illegal content and, where relevant, illegal activity and content and activity that is harmful to children.

1.27 The Act makes the Codes a "safe harbour" meaning service providers which implement all applicable measures in a Code are deemed compliant with the duty to which the Code relates.[18]

1.28 Service providers may choose to comply with their safety duties by taking "alternative measures". In doing this, service providers must keep a record of what they have done and explain how the relevant safety duties have been met. The Act provides that, if taking alternative measures, service providers should have regard to the importance of users' rights to freedom of expression and the right to privacy. We set out what providers should record when taking alternative measures in our Record-Keeping and Review Guidance.[19]

Online safety principles and objectives

1.29 Schedule 4 of the Act contains a set of General Principles and Online Safety Objectives for Ofcom. When preparing Codes, Ofcom must consider the appropriateness of the measures we recommend to different kinds and sizes of services and to providers of differing sizes and capacities.[20] We must also have regard to the following principles:

a) providers must be able to understand which measures apply to their services;
b) the measures must be sufficiently clear, and at a sufficiently detailed level, that providers understand what they entail in practice;
c) the measures must be proportionate and technically feasible; and
d) the measures must be proportionate to our assessment of the risk of harm presented by the service of that kind or size.

[17] Part 3, 4, 5 and 6 of the 2023 Act.
[18] Section 49 of the 2023 Act.
[19] Ofcom (2024), Record-Keeping and Review Guidance.
[20] Schedule 4 of the 2023 Act.

1.30 We must also ensure that the measures described in the Codes are compatible with the pursuit of the list of online safety objectives set out in Schedule 4, and that we include measures relating to each of the areas specified in sections 10(4) and 27(4) (concerning illegal content) and sections 12(8) and 29(4) (concerning content harmful to children).[21]

1.31 Under the 2003 Act, we are also required to conduct impact assessments when preparing a Code or amendment to a Code, including an assessment of the impact on small and micro businesses.[22]

Human rights

1.32 It is unlawful for Ofcom to act in a way which is incompatible with the European Convention on Human Rights (ECHR).[23]

1.33 Of particular relevance to our functions under the Act are the right to freedom of expression (Article 10 ECHR) and the right to privacy (Article 8 ECHR).[24] In formulating our decisions in this statement, we have identified the potential for interference with ECHR rights and considered whether any such interference is proportionate.

1.34 The right to freedom of expression includes the freedom to hold opinions and to receive and impart information and ideas without interference by public authority. Article 10(2) of the ECHR states that this right may be restricted in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.

1.35 Article 8(1) of the ECHR states that everyone has the right to respect for private and family life, home and correspondence. Article 8(2) sets out that public authorities must not interfere with the exercise of this right unless necessary in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

1.36 Other ECHR rights which may be relevant to our functions under the Act are the right to freedom of thought, conscience and religion (Article 9 ECHR) and the right to freedom of assembly and association (Article 11 ECHR).

---

[21] For user-to-user services, these areas include design of functionalities, algorithms, and other features; policies on user access to the service or to particular content present on the service, including blocking users from accessing the service or particular content; and content moderation, including taking down content. For search services, these areas include design of functionalities, algorithms and other features relating to the search engine; and content prioritisation. We discuss the areas specified in sections 10(4), 12(8), 27(4) and 29(4) of the Act in more detail in annex 3: Legal framework.

[22] Section 7 of the 2003 Act, as amended by section 93 of the Act.

[23] Section 6 of the Human Rights Act 1998.

[24] For service providers, the right to peaceful enjoyment of their possessions is also relevant (Article 1 of the First Protocol: "No one shall be deprived of his possessions except in the public interest and subject to the conditions provided for by law and by the general principles of international law"). However, this in no way impairs the right of the UK to enforce such laws as it deems necessary to control the use of property in accordance with the general interest. We consider the Act to be such a law and we consider our impact assessment as a whole to demonstrate that the measures are proportionate.

1.37 These are qualified rights. However, the need for any interference with these rights must be construed strictly and established convincingly. Any interference must be prescribed by or in accordance with the law, pursue a legitimate aim, and be necessary in a democratic society – in other words, it must be proportionate to the legitimate aim pursued and corresponding to a pressing social need.

1.38 In considering whether impacts on these rights are proportionate, our starting point is to recognise that the UK Parliament has determined that providers of regulated services must take proportionate measures to protect users from illegal content and, where relevant, address the risks of the service being used for the commission or facilitation of priority offences. We therefore start from the position that UK users should be protected from the harms set out in the Act and place weight on all the specific evidence of harm set out in our Illegal Harms Register of Risks (Illegal Harms Register). A substantial public interest exists in these outcomes. Overall, we have sought to strike a fair balance between securing adequate protections for users from harm (and their human rights in respect of this) and the ECHR rights of users, other interested persons (including for example, persons who host websites or who may be featured in content on regulated services or whose content might be on those services regardless of whether or not they are service users) and service providers, as relevant.

1.39 Rules on data protection are also relevant to the right to private life. Service providers will need to comply as appropriate with the data protection principles, such as accuracy, fairness, transparency, and data minimisation, as well as individuals' rights to appropriate safeguards when subject to automated decision-making. To the extent relevant to the right to private life, these are considered in the rights assessment section of the relevant chapters.

1.40 More generally, data protection law is likely to be an important consideration in the provision of regulated online services. The Information Commission's Office (ICO)[25] is the regulatory authority on data protection and information rights and providers should have regard to relevant ICO guidance. We have sought to signpost to relevant guidance documents where appropriate but providers should always check the ICO's website for complete and up to date information.

Equality and Welsh Language Impact Assessment

1.41 We have considered the equality impacts of the Codes and guidance set out in this statement, detailing our understanding of any particular impacts on protected groups in the UK.

1.42 Where relevant, and to the extent we have discretion to do so in the exercise of our functions, we have considered the potential impacts on opportunities to use the Welsh language and the need to treat the Welsh language no less favourably than English (in accordance with Welsh language standards).

1.43 We have set out our considerations on these matters in annex 2: Equality impact assessment and Welsh Language Assessment.

[25] The Information Commission's Office (ICO), formerly the Information Commissioner's Office, changed its name since the June 2025 Consultation. We will refer to them by their new name for the remainder of the statement.

The Deregulation Act 2015

1.44 The Growth Duty set out in the Deregulation Act 2015 requires Ofcom to have regard to the desirability of promoting economic growth when exercising our regulatory functions.[26] The Growth Duty has applied to our online safety functions since April 2026. This means that in our June 2025 Consultation, the Growth Duty was not explicitly referenced, although we considered the desirability of promoting economic growth in our assessment of the impacts of our measures. This statement contains explicit reference to the Growth Duty, and we have reassessed both individual measures in the 'Impacts and costs' section where appropriate, and in our combined impact assessment. See the chapter 13: Combined impact assessment for further information about the Growth Duty and our assessment.

UK Government decisions relevant to this statement

UK Government strategic priorities

1.45 In July 2025, the Secretary of State published the UK Government's strategic priorities for online safety.[27] We must have regard to a designated statement of strategic priorities for online safety when carrying out our online safety functions.

1.46 The July 2025 priorities were:

a) implementing safety by design to stop more harm occurring in the first place;
b) increasing transparency and accountability of online platforms;
c) maintaining regulatory agility to keep pace with changing technology and behaviour;
d) building an inclusive and resilient online society of well-informed users; and
e) supporting continued innovation in safety technologies.

1.47 We are addressing these priorities, in part, through the range of measures within this statement.[28]

UK Government social media ban

1.48 Earlier this year, the UK Government announced plans to introduce new legislation to restrict under-16s' access to social media and restrict their access to certain harmful functions on a wider range of services. The government also announced it would introduce default restrictions on persuasive design features, like infinite scroll, and overnight limits to help 16 and 17-year-olds have more control over their online experiences and help them manage their time on social media.[29] As the additional safety measures were proposed in June 2025, elements of the proposals have been reviewed in light of the announcement. These are addressed in each chapter.

[26] Section 108 of the Deregulation Act 2015.
[27] Secretary of State for Science, Innovation and Technology, 2025. Final Statement of Strategic Priorities for Online Safety - GOV.UK.
[28] Ofcom (2025), Letter to Government on the Statement of Strategic Priorities for Online Safety.
[29] UK Government (2026), Growing up in the online world: a national consultation - GOV.UK.; UK Government (2026), Social media to be banned for under-16s in landmark government move to give kids their childhood back - GOV.UK, UK Government, 2026. Growing up in the online world: government response (July 2026).

New priority offences

1.49 In December 2025, the government created two new 'priority offences' under the Online Safety Act: encouraging or assisting serious self-harm (the 'self-harm offence'), and cyberflashing. Both were previously non-priority offences under the Act. Providers have specific duties under the Act in relation to priority offences.

1.50 In our June 2026 New Priority Offences Statement, we determined that the offences of encouraging or assisting suicide and self-harm should be treated as one single 'kind' of illegal harm.[30] This means service providers will need to assess both suicide and self-harm, and assign one overall risk level for 'suicide and self-harm'. Measures that previously specifically applied to suicide content or services at medium or high risk of suicide will, once the revised Codes are in force, also apply to self-harm content and relevant services at medium or high risk of suicide and self-harm.

1.51 For the additional safety measures set out in this statement, we have decided that ICU D17 on livestreaming reporting and ICU E2 on recommender systems will apply to user-to-user services with relevant characteristics at medium or high risk of suicide and self-harm.[31] This change is reflected in the relevant chapters. We thank stakeholders who had requested this change in response to our original measures.[32]

Next steps

1.52 The Codes relating to the measures in both this statement and the June 2026 Statement will be laid in Parliament on 16 November 2026. The confirmed measures, draft Proactive Technology Guidance, draft Guidance on Human Review for Hash Matching and draft Guidance on Highly Effective Age Assurance for Part 3 Services will come into force in February 2027, subject to the UK Parliamentary process being completed. Providers will need to take the steps laid down in the Codes or use other effective measures to protect users (see paragraph 1.28).

[30] Ofcom (2026), June 2026 Statement, New priority offences serious self-harm and cyberflashing.
[31] We refer to these in the Illegal Content Codes as ICU D17 and ICU E2. Measures beginning with 'ICU' are included in the Illegal Content Codes of Practice or the Illegal Content Codes of Practice for user-to-user services: Proactive Technologies (unknown content) and measures beginning with PCU are included in the Protection of Children Code of Practice.
[32] Molly Rose Foundation response to the June 2025 Consultation, pp.3-16; Samaritans response to the June 2025 Additional Safety Measures Consultation, pp.1-2, p.7.

2. Livestreaming

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

What is this chapter about?

Livestreaming is a function where a user can create and distribute a live video that can be viewed by multiple other users. This chapter covers our full definition of livestreaming, explains the risks and benefits of it, and sets out our decisions on our proposed measures on reporting the risk of imminent physical harm on livestreams, the moderation of livestreams by humans, and protections for children using livestreaming functionality.

What decisions have we made?

We have decided that livestreaming services should provide users with a mechanism to report livestreams depicting an imminent risk of physical harm, including of animal cruelty.

We have decided not to proceed with our proposal that services should ensure human moderators are available whenever users can livestream; we have also decided to suspend work on our proposal to restrict certain functionalities during children's livestreams.

We are proceeding with the following measure:

Number in Codes	Recommended measure	Who should implement this
ICU D17	Livestreaming reporting mechanism for imminent physical harm	Providers of user-to-user services that: a) have one-to many livestreaming; and b) are at medium or high risk of one or more specified kinds of illegal harm.

Why have we made these decisions?

We consider that a dedicated reporting mechanism for imminent physical harm is a proportionate and effective way to help services identify and respond more quickly to serious illegal harms that can occur on livestreams. The measure will provide users with a clear means of reporting content that may involve an immediate risk of harm and support services in prioritising their response.

We have not proceeded with the proposed human moderation measure because we have concluded that existing content moderation measures within our Illegal Content Codes of Practice for user-to-user services already provide a proportionate and flexible basis for services to ensure moderation resourcing is appropriate to address the risks arising from livestreamed content.

We have decided to suspend work on the proposed children's livestreaming measure in light of the UK Government's announced plans, including restrictions on under-16s' access to social media and certain high-risk functionalities, including livestreaming. The Government's plans make the livestreaming proposals we had made unnecessary at this point in time.

Introduction

2.1 The Online Safety Act 2023 (the Act) requires providers of regulated user-to-user services to take certain steps to reduce the risk of harm to users from illegal content. The requirements include taking proportionate measures relating to the design or operation of a service to mitigate and manage the risks of harm to individuals (section 10(2) and (3)).

2.2 In our June 2025 Additional Safety Measures Consultation (June 2025 Consultation), we outlined two scenarios that we were particularly concerned about in relation to livestreaming:[33]

- Users creating livestreams to broadcast illegal content, in particular content where there is an imminent risk of physical harm. This includes terrorist, violent and hateful attacks, content which encourages or assists suicide or other livestreams of actual, imminent physical harm or violence, or the risk of it.
- Children being subject to a range of harms while livestreaming, enabled by viewers' ability to post comments and reactions, send gifts, and record the content of their livestreams.

2.3 We proposed three measures to protect against these scenarios, as follows:

a) services to enable users to report instances of imminent physical harm on livestreams (ICU D17)
b) services to have human content moderators available whenever users were able to livestream (ICU C16)
c) services to restrict certain functionalities when children were livestreaming to help prevent various forms of harm, most particularly child sexual exploitation and abuse (CSEA) (ICU F3).

2.4 In this chapter, we first set out an overview of our decisions relating to these measures. We then review our approach to livestreaming, and explain what it is and the factors that make it particularly risky.

2.5 Lastly, we review each of the three measures in turn, providing further detail on how we reached our final decisions.

Overview of our decisions

2.6 Having reviewed stakeholder feedback, and following the UK Government's plans to introduce new legislation, we have made the following decisions:

- Confirmation of measure ICU D17 as proposed in the June 2025 Consultation. Based on stakeholder feedback, we have made some small amendments to the measure to include animal cruelty in the harms in scope.
- Withdrawal of measure ICU C16 as proposed in our June 2025 Consultation. Having reviewed stakeholder feedback, we have decided not to proceed with this measure, as its effectiveness is uncertain and the existing Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes) already support proportionate and effective moderation of livestreaming content.[34]

[33] Ofcom (2025), June 2025 Additional Safety Measures Consultation, paragraphs 4.10-4.13.
[34] Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes).

- Withdrawal of measure ICU F3 as proposed in our June 2025 Consultation. This decision follows the UK Government's announcement that it intends to restrict under-16s' access to social media, and that it intends to restrict features including livestreaming for under-16s on a wider range of online services, including on gaming sites.

2.7 In addition to these changes, we have also refined the definition of "livestreaming" to clarify that, for the purpose of the measure in this statement, livestreaming relates to a user-to-user service functionality that allows a single user to simultaneously create and distribute live video depicting real-world events as they occur online, streaming content (such as live video content) to multiple or potentially unlimited other users in, or very close to, real time; this does not include private or closed communication services such as one-to-one or group video calls. See paragraphs 2.20-2.21 for further details.

2.8 We consider that other measures in this statement will also help to protect users, including those who livestream. For example, the reporting function discussed in this chapter, our measures on user sanctions and our extended default settings for child users will all deliver added protections for users' livestreams.[35]

Our approach to livestreaming

Definition of livestreaming

2.9 In our June 2025 Consultation, we stated that our proposed measures would apply to one-to-many livestreaming services. We defined this as a user-to-user service functionality that allows a single user to simultaneously create and broadcast online streaming content (such as live video content) to multiple other users in, or very close to, real time. We assessed that the principal means of conducting a livestream in this way is through a user livestreaming on a social media service, or on a dedicated livestreaming service. One-to-many livestreams may be visible to, and able to be accessed by, a large number of users. This may include a user's followers, friends or connections, or it may include all users or viewers of a service.[36]

2.10 We said that one-to-many livestreaming does not include instances when multiple users are streaming content with each other simultaneously, for example as part of a group chat or a multiplayer game.

Feedback on our definition of livestreaming

2.11 Several stakeholders commented on our definition of livestreaming. This included suggestions to extend the scope of the measures to capture other forms of livestreaming (such as many-to-many and/or co-hosted livestreams, where multiple people broadcast content simultaneously), which some considered to also present risks to children in particular.[37]

[35] See chapters 3: User sanctions and chapter 6: Increasing the effectiveness of safety defaults and supportive messaging.
[36] Ofcom (2025), June 2025 Additional Safety Measures Consultation, paragraphs 4.1-4.9.
[37] 5Rights Foundation response to the June 2025 Additional Safety Measures Consultation (June 2025 Consultation), pp.5-6; Canadian Centre for Child Protection (C3P) response to the June 2025 Consultation, pp.4-5; International Justice Mission response to the June 2025 Consultation, pp.2-8; Internet Watch Foundation (IWF) response to the June 2025 Consultation, pp.4-5; Marie Collins Foundation response to the June 2025 Consultation, pp.4-5; National Society for the Prevention of Cruelty to Children (NSPCC) response to

2.12 The majority of this feedback related to our children's livestreaming measure (ICU F3). We are no longer proceeding with this measure following the UK Government's announcement of plans to restrict under-16s' access to social media and its intention to restrict features including livestreaming and stranger communication with children for under-16s on a wider range of online services, including on gaming sites. The government has also announced services will need to restrict these features (livestreaming and communicating with strangers) by default for 16 and 17 year olds.

2.13 We also received requests to clarify whether functions such as one-to-one livestreams and/or live video calls were in scope of our definition; some stakeholders expressed support for these functions (especially one-to-one livestreams) being in scope, while others argued that they should be kept outside of the scope of the measures.[38]

Application of the measures to gaming services

2.14 We received feedback on the application of the proposed measures to gaming services.

2.15 The Association for UK Interactive Entertainment (Ukie) raised several points as to why it did not consider the proposed measures to be suitable for gaming services. It said that livestreaming in video games is a distinct, low risk use case and that livestreaming in games is usually limited, and services already embed robust safety measures therefore a one-size fits all approach could harm innovation. It also said that the proposed measures were too prescriptive and impractical for gaming environments and alternative safety by design measures would be more proportionate. In relation to ICU C16 it said requirements for human moderation were excessive, given existing controls and restrictive nature of communication, and do not reflect how livestreaming works in games. It also stated that the existence of safeguards – such as real-time moderation, filtering, reporting tools, delay buffers, and automated risk detection meet the intended outcomes of our measures and that available evidence does not justify the proposed level of regulatory intervention.[39]

2.16 As explained in the section titled 'Overview of our decisions', we have decided not to proceed with measures ICU C16 and ICU F3.

2.17 With regards to measure ICU D17 (reporting of imminent physical harm) we wish to clarify that:

a) where a service makes available a functionality capable of sharing real-world video in real time, the service will fall within scope of our measure, regardless of whether users routinely use that functionality;
b) where a service enables users to share real-world video in real time as part of a broader experience (for example, through a webcam overlay accompanying gameplay), it will be considered capable of depicting real-world content and will therefore fall within scope of our measure; and

---

the June 2025 Consultation, pp.10-12; Online Safety Network (new name) response to the June 2025 Consultation, pp.9-10. The Online Safety Network, formerly the Online Safety Act Network, changed its name since the June 2025 Consultation. We will refer to them by their new name for the remainder of the statement.; South West Grid for Learning (SWGfL) response to the June 2025 Consultation, p.7; [✂].

[38] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.2; Canadian Centre for Child Protection (C3P) response to the June 2025 Consultation, pp.4-5; Computer and Communications Industry Association (CCIA) response to the June 2025 Consultation, p.4; International Justice Mission response to the June 2025 Consultation, pp.2-8; Meta Platforms Inc. (Meta) response to the June 2025 Consultation, p.3; [✂]; [✂]; National Police Chiefs' Council (NPCC) response to the June 2025 Consultation, pp.2-3; [✂].

[39] Association for UK Interactive Entertainment (Ukie) response to the June 2025 Consultation, pp 3, 8, 12-13.

c) services such as online games that are limited to depicting gameplay or other virtual environments, and do not provide any functionality to capture or transmit real-world imagery, will not fall within scope, unless they integrate features that allow real-world events to be livestreamed in real time.

2.18 The measure does not apply to services posing a low risk of the identified harms, regardless of service type.

2.19 Where gameplay is livestreamed via a separate service, that livestreaming service is in scope, not the underlying game – unless the game itself offers real-time livestreaming of real-world imagery.

Our final definition

2.20 Following stakeholder feedback, we have refined our definition of one-to-many livestreaming for the purposes of ICU D17. We define this as: "A user-to-user service functionality that allows a single user to simultaneously create and distribute live video depicting real-world events as they occur online, streaming content (such as live video content) to multiple or potentially unlimited other users in, or very close to, real time. This excludes content which is communicated privately."

2.21 This definition is intended to focus the measure on livestreaming capabilities that could realistically depict real-world events as they unfold and clarifies that it excludes private or closed communication services such as one-to-one or group video calls. This approach ensures proportionality, supports innovation, and targets obligations at services presenting the relevant risks.

Benefits and risks of harm associated with livestreaming

2.22 Evidence suggests that people livestream in large numbers. When surveyed, over a quarter (28%) of children aged 8-17 told us that they had broadcast their own livestreams.[40] Among those aged 18 or over, 21% of men and 13% of women told us that they had broadcast livestream content.[41] [42]

2.23 Evidence also suggests that livestreaming is used for a variety of different purposes, with an array of associated benefits and harms.[43] [44] [45] The stakeholder feedback we have received corroborates this view. We expand on the evidence of benefits and risks of harm, as well as on the stakeholder feedback, in the following section.

---

[40] Ofcom (2026), Children's usage of livestreaming (Livestreaming report 3), p.10.
[41] Ofcom (2025), Adults' Media Use and Attitudes Report.
[42] It should be noted that no specific definition of livestreaming was provided to respondents to this survey, and so the figures are likely to encapsulate all types of livestreaming. This could potentially include one-to-one video calls. For this reason, these figures are not directly comparable to the aforementioned children's study, which did include a specific livestreaming definition.
[43] Ofcom (2026), Children who create and view livestreams.
[44] Huang, Z., Mou, J., Benyoucef, M. and Kim, J., 2023. Live Streaming: Its Relevant Concepts and Literature Review.
[45] Anecdotal evidence of the purposes and benefits of livestreaming for children has been gathered from the following sources: Internet Matters, 2025. What is live streaming & vlogging?; NSPCC, 2025. Livestreaming and online video apps.

Benefits associated with livestreaming

2.24 In the June 2025 Consultation, we set out that livestreaming is used for many social, entertainment and commerce purposes. Other uses may include citizen journalism, social activism and real-time crisis reporting. We considered that livestreaming allows for the development of, and engagement with, communities of common interest. We noted that identified benefits to livestream broadcasters include engagement with and receiving support from others, immediacy of interaction and feedback, development of creative, entrepreneurial, public speaking and communication skills and confidence, and financial opportunities.[46]

2.25 In responding to the June 2025 Consultation, several stakeholders from across civil society, industry and government agreed that livestreaming delivers multiple benefits to users, including children, with some providing further evidence and views regarding those benefits.[47]

2.26 Since our June 2025 Consultation, we have conducted further research to support our understanding of children's and young people's experiences of livestreaming. This includes undertaking qualitative and quantitative research studies to examine the motivations, experiences, benefits, and risks of livestreaming to children, both as creators and viewers of this form of content.[48] We also conducted a Behavioural Audit of Livestreaming Platforms to examine livestreaming functionalities on popular services and how they are likely to influence user behaviour on those services.[49]

2.27 This research presented a complex picture of children's engagement with livestreaming. Over a quarter (28%) of children aged between 8-17 said they had ever created a livestream themselves, [50] with a high proportion (77%) of those children doing so recently and regularly (once a week or more often).[51] Regular livestreamers from our qualitative study told us that the live nature of the livestreaming format and its associated interaction features (such as the chat function) had advantages as it felt more authentic and, compared to recorded video, provided a stronger sense of connection with others, however these features also posed risks.[52]

---

[46] Ofcom (2025), June 2025 Additional Safety Measures Consultation, paragraphs 4.4-4.6.
[47] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.3; CCIA response to the June 2025 Consultation, pp.2-3; Fish, A., response to the June 2025 Consultation, pp.1-3; Google LLC. response to the June 2025 Consultation, p.8; Internet Matters response to the June 2025 Consultation, p.3; [✂]; Online Dating and Discovery Association (ODDA) response to the June 2025 Consultation, p.2; Open Rights Group response to the June 2025 Consultation, pp.2-4; Thompson, A. response to the June 2025 Consultation, p.2; [✂]; Twitch Interactive Inc. (Twitch) response to the June 2025 Consultation, pp.2-5; Yoti response to the June 2025 Consultation, pp.2-3.
[48] Ofcom (2026), Livestreaming: The perspective of children and young people who create livestreams; Ofcom, 2026. Research among children and young people who view livestreams; Ofcom (2026), Children's usage of livestreaming (Livestreaming report 3).
[49] Ofcom (2026), Behavioural Audit of Livestreaming Services: Key Findings Report.
[50] Ofcom (2026), Children's usage of livestreaming (Livestreaming report 3), p.2.
[51] Ofcom (2026), Children's usage of livestreaming (Livestreaming report 3), p.10.
[52] Ofcom (2026), Livestreaming: The perspective of children and young people who create livestreams, p.12-15.

Risks of harm associated with livestreaming

2.28 In our June 2025 Consultation, we also set out that livestreaming can pose multiple risks to users.[53] Many of these are linked to the instantaneous and transitory nature of the content and interactions that take place, and challenges in moderating livestreamed content. We noted that livestreaming and its associated functionalities such as the ability to comment, send reactions and gifts, and screen record content, present particular risks to children, and provided evidence of harm occurring as a result, particularly in relation to grooming and coercion into producing child sexual abuse material (CSAM).[54]

2.29 Multiple stakeholders supported our assessment of the risks and harms that livestreaming presents to users, including to children.[55] 5Rights Foundation said that extensive evidence in the Children's Register of Risks (Children's Register), combined with the use of livestreaming features, created a clear imperative to act.[56] The Online Safety Network argued that our evidence base surrounding livestreaming was already well established, particularly from our previous November 2023 Illegal Harms Consultation (November 2023 Consultation) and the Children's Register.[57]

2.30 Some stakeholders however disagreed with our assessment and suggested that the majority of evidence we had cited was not specific to livestreaming and/or often related to gaming.[58] Other stakeholders suggested additional harms that we could consider as part of the livestreaming measures, including AI or gendered harms, such as sexual solicitation, coercion and misogynistic abuse.[59]

2.31 While, as noted in paragraph 2.6, we are not proceeding with the children's livestreaming and human moderation measures, we did not receive any evidence from stakeholders that materially changed our assessment of the risks of harm that measure ICU F3 in particular seeks to address. We will consider responses and evidence provided on both the benefits and risks of livestreaming to children as part of any future work.

2.32 In addition to setting out the possible benefits of livestreaming as detailed in paragraphs 2.24-2.27, the research we have conducted since the June 2025 Consultation also

[53] Ofcom (2025), June 2025 Additional Safety Measures Consultation, paragraph 4.6.
[54] Ofcom (2025), June 2025 Additional Safety Measures Consultation, paragraphs 4.36-4.75.
[55] 5Rights response to the June 2025 Consultation, p.3; A member of the Philippine Survivor Network response to the June 2025 Consultation, pp.2-3; C3P response to the June 2025 Consultation, pp.1-3; Children's Commissioner's Office (CCO) response to the June 2025 Consultation, pp.1-2; Christian Action Research and Education (CARE) response to the June 2025 Consultation, pp.1-2; Common Sense Media response to the June 2025 Consultation, pp.2-4; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, pp.2-3; Internet Matters response to the June 2025 Consultation, pp.2-3; IWF response to the June 2025 Consultation, pp.2-3; Molly Rose Foundation response to the June 2025 Consultation, pp.10-13; NPCC response to the June 2025 Consultation, p.2; Online Safety Network response to the June 2025 Consultation, pp.2-4; SWGfL response to the June 2025 Consultation, p.8 [✂]; UK Finance response to the June 2025 Consultation, p.3.
[56] 5Rights response to the June 2025 Consultation, pp.2-3.
[57] Online Safety Network response to the June 2025 Consultation, pp.2-4.
[58] [✂]; Twitch response to the June 2025 Consultation, pp.10-11; [✂].
[59] British and Irish Law Education and Technology Association (BILETA) response to the June 2025 Consultation, pp.2-3; Refuge response to the June 2025 Consultation, pp.2-4.

highlights ways in which the real-time nature of livestreaming could present risks more directly.[60]

2.33 The children we spoke to as part of our research talked about the limited control of and ability to predict interactions, instances of unwanted contact, and pressure to perform or remain visible.[61] Consistent with this, the behavioural audit also found that livestreams were typically public by default. In some cases, users were encouraged when ending their livestream to continue streaming through pop-ups on screen, or goal-oriented statistics of how their livestream had performed, and how close they were to reaching a new level of reward on the service. These design features actively encourage user visibility through livestreams.[62]

2.34 Some younger creators in our research (those aged 15 and under) were less aware that anyone could access their livestreams and the potential risks relating to this, such as other users being dishonest about their identity.[63] This lack of understanding of online risk is also described in our 'Child development ages, stages and online behaviour' report, which indicates that younger children tend to have a limited ability to critically evaluate information or others' intentions and may overestimate their ability to cope with challenges online.[64]

2.35 Furthermore, in several instances, children in our qualitative sample said they were creating livestreams on platforms whose terms of service state that children of their age are not permitted to livestream.[65] Findings from the quantitative research, while based on claimed behaviour, appear to suggest that this occurs more widely.[66] This is corroborated by our Behavioural Audit, where researcher accounts were not required to age assure or on any services that were subject of the audit, demonstrating the ease with which children could create accounts with livestreaming functionality, despite this being against the terms of service.[67] This is particularly concerning given that services themselves have set a minimum age to their service or this functionality, which indicates that they consider livestreaming inappropriate for children of certain ages.

Conclusion

2.36 In summary, the evidence and feedback discussed did not materially change our assessment of the risks of harm that our measures seek to address. Our assessment of the harmful scenarios that we were particularly concerned about in relation to livestreaming, as outlined in paragraph 2.2 above, has not changed following the June 2025 Consultation. As detailed in paragraphs 2.62-2.64, we have also added animal cruelty to the physical harm scenarios. We maintain our assessment that urgent risks to life can arise from livestreaming, including the risks described in paragraph 2.2.

---

[60] Ofcom (2026), Livestreaming: The perspective of children and young people who create livestreams. Ofcom, 2026. Research among children and young people who view livestreams; Ofcom (2026), Children's usage of livestreaming (Livestreaming report 3).
[61] Ofcom (2026), Livestreaming: The perspective of children and young people who create livestreams, p.41.
[62] Ofcom (2026), Behavioural Audit of Livestreaming Services: Key Findings Report, p.10-11.
[63] Ofcom (2026), Livestreaming: The perspective of children and young people who create livestreams, p.43.
[64] Ofcom (2025) Child development ages, stages and online behaviour, p.6.
[65] Ofcom (2026), Livestreaming: The perspective of children and young people who create livestreams, p.20.
[66] Ofcom (2026), Children's usage of livestreaming (Livestreaming report 3), p.14.
[67] Ofcom (2026), Behavioural Audit of Livestreaming Services: Key Findings Report, p.5-6.

2.37 We expect that our measure on including a reporting category for imminent physical harm is a step in the right direction. However, we are conscious that more would need to be done to protect users from the risks this functionality holds. We will be closely following the UK Government's forthcoming legislation, which includes restrictions for livestreaming, to decide on next steps.

Measure on reporting mechanism for imminent physical harm

2.38 In the June 2025 Consultation, we proposed that livestreaming providers should have a mechanism to enable users to report livestreams containing content that depicts a risk of imminent physical harm.

2.39 We said that this measure would apply to all user-to-user services offering one-to-many livestreaming at medium or high risk of terrorism content, grooming or image-based child CSAM, encouraging or assisting suicide, hate, or harassment, stalking, threats, and abuse offences. This included services whose primary functionality was livestreaming and those where it was an additional feature.

2.40 We said that this measure would help users quickly and easily report livestreams that contain suspected illegal content that includes physical harm – for example, a live terrorist attack, CSEA content, or content relating to a potential suicide attempt.

2.41 By implementing this measure, we explained that services would be provided with specific signposts to those livestreams that may contain this kind of illegal content. We said that this would enable service providers to identify content that depicts an imminent risk of physical harm more quickly and easily than would otherwise be the case and take appropriate action.

Our decision

2.42 Based on the responses received, we have decided to broadly confirm the measure we proposed in the June 2025 Consultation. We have decided to extend the harms in scope to include animal cruelty following feedback received (see paragraphs 2.62-2.64 for our reasoning). We also received responses asking for clarification on which gaming services are in scope. We provided this clarification in paragraphs 2.14-2.19.

2.43 This measure will apply to all user-to-user services offering one-to-many livestreaming at medium or high risk of one or more of the following harms: terrorism content, grooming or image-based CSAM, encouraging or assisting suicide and self-harm, hate, or harassment, stalking, threats and abuse and animal cruelty offences.[68] This includes services whose main function is to provide livestreaming and those where it is a supplementary or secondary feature.

2.44 The full text of the measure and the services in scope can be found in our Illegal Content user-to-user Codes of Practice for Terrorism, CSEA and for other duties and is referred to as ICU D17.

[68] As explained chapter 1: Introduction, our duties and navigating the Statement, suicide and self-harm is now considered a single kind of illegal harm across this statement.

How this measure works

2.45 Under this measure, providers should ensure their reporting functions or tools have a category that allows the user to report instances of imminent physical harm on livestreams.

2.46 Providers may have other reporting categories, which may be useful for reporting other types of suspected illegal content or content harmful to children, such as "dangerous activities and challenges", "violent content", "child safety", "terrorism content" or "suicide and self-harm". However, we consider that these examples are not specific enough to alert services to suspected imminent physical harm on livestreams.

2.47 Providers can choose to use alternative wording for this category that has the same meaning as "imminent physical harm" and ensures the same intended outcome of this measure. The design of the reporting mechanism/naming of the category sits with the provider and some providers may already have a tool to this effect in place for users.

2.48 This measure applies in respect to complaints made by UK users and/or interested persons in the UK. Where a UK user can watch a livestream, the user should have the reporting category available to them. The reporting category should be available on all livestreams regardless of where it originates.

2.49 We would expect providers in scope of measures ICU C5 (Content moderation – prioritisation) and ICU D7 (Content moderation – training and materials) to deal with these reports as part of their prioritisation policy once they have received them.

Our reasoning

Benefits and effectiveness

2.50 As set out in paragraphs 2.28-2.35, livestreaming functionality can pose multiple risks to users due to the perceived real-time and short-lived nature of the content.

2.51 Given the real-time nature of livestreams, reporting is an important mechanism and complementary safeguard alongside existing proactive and reactive systems adopted by services to become aware of suspected illegal content.

2.52 Livestreaming creates a particular risk of immediate physical harm because of the immediate broadcast nature of the functionality, with little or no opportunity for preventative moderation. Real-time audience interaction and the potential for content to spread quickly can further escalate harm, though some services do include moderation within the chat interactions.[69]

2.53 In our view, it is important that a mechanism exists to allow users to report a livestream that depicts suspected imminent physical harm. This means that services can be provided with a specific signpost to those livestreams that may contain the illegal harms in scope of this measure.

2.54 As explained in our June 2025 Consultation, effective reporting functions can reduce the length of time that a livestream is online after harmful content has been reported, which in turn can reduce harm to users viewing the livestream at the time.

2.55 Cutting a livestream short can also limit the perceived 'value' of recording the livestream by perpetrators, making them less likely to become viral or invoke 'copy-cat' actions. Evidence

[69] Ofcom (2026), Behavioural Audit of Livestreaming Services: Key Findings Report, p.9.

from previous incidents shows how the continued availability of livestreamed harmful content can significantly amplify its impact.[70] In contrast, earlier interruption of a livestream limits the scale and longevity of harmful recordings.[71]

2.56 Evidence also suggests that perpetrators of serious illegal harm may use livestreaming to maximise real-time visibility, symbolic impact, and the later circulation of recordings. Public investigations and academic analysis show that the scale, persistence, and amplification of livestreamed content can contribute to its wider harm and influence.[72]

2.57 In our view, the availability on a livestreaming functionality of a reporting category for imminent physical harm would enable users to quickly and easily report livestreams with suspected illegal content such as live terror attacks, content encouraging or assisting suicide and self-harm, CSEA and hate, threats, abuse and harassment and animal cruelty.[73] This measure could also enable more rapid interruption of livestreams and swift removal of content, thus reducing the reach and longevity of such material, particularly where intervention occurs before an incident concludes.

2.58 The use of this reporting category by the users watching livestreams could enable overall faster moderation of suspected content where there is a report of imminent physical harm threat to life. This will enable livestreaming services to reduce further harm to users through the creation and onward dissemination of illegal content.

2.59 For these reasons, we have concluded that the measure would be an effective way of combatting the risks associated with livestreaming and would deliver benefits.

2.60 Some stakeholders agreed with the value of a reporting measure, noting it could prompt moderators and enable services to prioritise content and could enable users to alert

---

[70] Following the 2019 Christchurch terrorist attack, footage of the attack remained available online for the duration of the livestream and was later recorded and shared at scale. In the 24 hours that followed, Facebook removed 1.5 million videos of the attack, with a further 1.2 million attempts blocked at upload. Despite these efforts, researchers at the Counter Extremism Project identified at least 14 different websites hosting footage of the attack a year later. This illustrates how delays in interrupting livestreams can increase the volume, longevity, and circulation of recordings, underscoring the importance of swift action to limit harm and wider replication. Office of the New York State, 2022. Buffalo Shooting Online Platform Investigative Report.

[71] On 14 May 2022, an attack was carried out in Buffalo, New York, which resulted in the death of ten individuals and the wounding of three others. The attack was livestreamed online, and versions of the footage were disseminated on multiple online services, potentially exposing UK users to content related to terrorism. Ofcom, 2023. BitChute improves its safety measures following engagement with Ofcom - Ofcom.
The New York State Attorney General found that Twitch disabled the Buffalo livestream within approximately two minutes, an improvement on the Christchurch attack in 2019. While Buffalo footage still spread online, in part due to a small service acting slowly to remove a video of the livestream, it did not show the same degree of long-term circulation, supporting the role of rapid intervention in reducing wider amplification. Office of the New York State, 2022. Buffalo Shooting Online Platform Investigative Report.

[72] Global Internet Forum to Counter Terrorism (GIFCT), 2021. Technical Approaches Working Group: Gap Analysis and Recommendations; Gorwa, R.,2024. After Christchurch: Diverging Regulatory Responses in New Zealand and Australia, Oxford Academic; New York State Office of the Attorney General, 2022. Investigative Report on the Role of Online Platforms in the Tragic Mass Shooting in Buffalo on May 14, 2022.
Zornetta, A. and Pohland, I., 2022. Legal and Technical Trade-offs in the Content Moderation of Terrorist Live-Streaming, International Journal of Law and Information Technology.

[73] We have decided to extend the harms in scope to include animal cruelty following feedback received (see paragraph 2.62-2.64 for our reasoning).

services to illegal content that may go undetected, or agreed with the approach in general.[74] [75]

2.61 While stakeholder responses have not led us to fundamentally revise our assessment of the benefits of this measure, several stakeholders provided feedback or requested clarity on specific points of detail which could affect its efficacy. Some stakeholders suggested that the proposed reporting measure was "unnecessary" or duplicative of existing reporting tools, or that it was overly "prescriptive".[76] Others requested greater clarity on how the measure would operate in practice.[77] [78] We review this along with other feedback in the following paragraphs.

Inclusion of animal cruelty

2.62 The Social Media Animal Cruelty Coalition (SMACC) and Digital Resilience in Education (Welsh Government) argued that animal cruelty should be added to the measure, citing evidence relating to the link between animal and human violence.[79]

2.63 In our Illegal Harms Register of Risks (Illegal Harms Register), we explain that livestreaming can be used to encourage the physical abuse of animals by users in real time, particularly through comments.[80] Our Risk Profiles for Illegal Harms similarly lists animal cruelty as an important risk factor in livestreaming.[81]

2.64 Having considered the stakeholder responses and noting that livestreaming is a risk factor for animal cruelty, we have decided to add the animal cruelty offence to the scope of harms.

Actions providers take in response to the report

2.65 Some stakeholders said that the proposed reporting measure did not contain any detail on follow-up action for providers and called for us to add to the measure. The respondents indicated that we should include that providers should act promptly on (or prioritise the

---

[74] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.6; Advisory Committee for Northern Ireland response to the June 2025 Consultation, p.1; BILETA response to the June 2025 Consultation, p.4; C3P response to the June 2025 Consultation, p.4; CARE response to the June 2025 Consultation, p.2; CCO response to the June 2025 Consultation, p.4; Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p.3; Department for Justice Northern Ireland response to the June 2025 Consultation, p.2; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.4; Domestic Abuse Commissioner response to the June 2025 Consultation, p.1; Evangelical Alliance response to the June 2025 Consultation, p.2; ICO response to the June 2025 Consultation, p.26; [✂]; Internet Matters response to the June 2025 Consultation, pp.4-5; Microsoft response to the June 2025 Consultation, p.2.; Name Withheld 4 response to the June 2025 Consultation, p.3; National Center for Missing & Exploited Children (NCMEC) response to the June 2025 Consultation, p.2; [✂]; Online Responsibility Network response to the June 2025 Consultation, p.1; Samaritans response to the June 2025 Consultation, p.1; Scottish Government response to the June 2025 Consultation, p.2; [✂].

[75] Stakeholders who expressed support for the measure (with some caveats) included: A member of the Philippine Survivor Network response to the June 2025 Consultation, p.5; Open Rights Group response to the June 2025 Consultation, p.4; Samaritans response to the June 2025 Consultation, p.2; [✂]; [✂].

[76] Name Withheld 22 response to the June 2025 Consultation, p.2; [✂].

[77] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.5; [✂].

[78] We set out further details on stakeholders who supported and disagreed with the measure in the relevant chapter in annex 1: Further stakeholder responses.

[79] Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.4; Social Media Animal Cruelty Coalition (SMACC) response to the June 2025 Consultation, p.4.

[80] Ofcom (2024), Illegal Harms Register of Risks (Illegal Harms Register), p.344.

[81] Ofcom (2024), Risk Assessment Guidance and Risk Profiles, p.38.

report of) imminent physical harm,[82] with Samaritans emphasising the need for a wider range of reporting categories and to prioritise suicide and self-harm content.[83]

2.66 Stakeholders highlighted that the measure lacked clarity on how providers should respond to reports, including expectations around the speed of action, the need for rapid human review and escalation where appropriate, and greater transparency – particularly providing users with feedback on outcomes – which they considered important for both effective enforcement and building user trust.[84]

Prompt action following reporting

2.67 We agree that prompt action is important where services receive reports of livestreams depicting imminent physical harm. However, our existing Illegal Content Codes already address this. In particular, the illegal content measures recommend providers to take swift action on illegal content of which they are aware, including priority illegal content (ICU C2). The illegal harms content moderation measures further set expectations around prioritisation (ICU C5), supported by appropriate performance targets (ICU C4) and resourcing (ICU C6), that apply to large services or multi-risk services.

2.68 In this context, we do not consider it necessary to introduce added provisions on content removal. Instead, the reporting measure should work alongside, and be supported by, the existing reporting and complaints and content moderation measures in the Illegal Content Codes.

Accessible reporting function

2.69 Our behavioural audit found that reporting tools are often difficult to access, requiring users to navigate away from the livestream and through multiple steps.[85] Given the real-time nature of livestreams, the availability of fast reporting is essential.

2.70 The aim of this measure is to make it straightforward for users to report livestreams depicting imminent physical harm. This removes the need for users to determine the specific illegal harm category (such as terrorism or hate) in real time, which may be difficult where risks are rapidly unfolding.

2.71 The provision of a clear and accessible reporting option to users enables services to identify high-risk content quickly and apply existing moderation processes to take prompt action in line with the duties.

2.72 In addition, our existing reporting and complaints measures already support the broader aim of accessible reporting and efficient handling of complaints. Measure ICU D7 (Appropriate action for relevant complaints about suspected illegal content) recommends that services take appropriate action in response to complaints about suspected illegal content. The measure sets out that complaints should be reviewed in line with content moderation measure ICU C1 unless identified as manifestly unfounded. Taken together,

---

[82] 5Rights Foundation response to the June 2025 Consultation, p.8; Molly Rose Foundation response to the June 2025 Consultation, p.9; Online Safety Network response to the June 2025 Consultation, p.7; SMACC response to the June 2025 Consultation, p.6.
[83] Samaritans response to the June 2025 Consultation, pp.1-2.
[84] 5Rights Foundation response to the June 2025 Consultation, p.7; Molly Rose Foundation response to the June 2025 Consultation, pp.9, 12; Online Safety Network response to the June 2025 Consultation, p.7; Samaritans response to the June 2025 Consultation, pp.2-3; SMACC response to the June 2025 Consultation, p.6.
[85] Ofcom (2026), Behavioural Audit of Livestreaming Services: Key Findings Report, p.9.

these measures provide sufficient clarity to ensure prompt and effective responses to reports of imminent physical harm.

Impacts and costs

2.73 Our June 2025 Consultation explained that implementation costs for a reporting category would include those related to design changes, engineering costs and further updates and refinement. These were relatively broad estimates that included the creation of prioritisation systems that services may develop to quickly respond to reports of imminent physical harm.

Direct costs

2.74 Our June 2025 Consultation explained that costs were higher estimates and would likely vary depending on the complexity of the service, as well as the existing set-up of reporting on the service.

2.75 The British and Irish Law Education Technology Association (BILETA) described our reporting category costs as "modest when measured against the anticipated safety benefits".[86] No consultation respondents provided evidence which gives us grounds for concluding that our initial cost estimates for this measure were incorrect. Our assessment of the direct costs for the reporting category therefore remains unchanged from our position in our June 2025 Consultation.

2.76 We have though updated our quantified estimates from the consultation in line with the latest Office for National Statistics (ONS) salary data.[87] Due to this update, we now estimate that in-scope services setting up a reporting category would incur a one-off implementation cost between £22,000 and £75,000, as well as ongoing annual maintenance costs ranging between £5,000 and £19,000.[88] This cost comprises the software engineering, legal and policy resource needed to add the category and create a prioritisation and review system for potential immediate physical harm. If services already have a prioritisation and review system, we estimate that solely adding the reporting category for imminent physical harm would cost £900 to £3,000.[89]

2.77 Other respondents did not specifically comment on cost estimates for our reporting category and were concerned more broadly with the level of costs across the combined reporting and content moderation measures.[90] We expect these concerns to diminish with our decision to remove the human moderation measure (see paragraphs 2.105-2.139).

Indirect impacts

2.78 Multiple stakeholders stated that our measures would likely have a larger relative impact on smaller services, with the potential for dissuading market entry and risking market consolidation.[91] In addition, two industry stakeholders discussed a negative impact on

---

[86] BILETA response to the June 2025 Consultation, p.4.
[87] See annex 5: Further detail on economic assumptions and analysis.
[88] This is based on two to three-and-a-half months of software engineering time with an equal amount of time from other professional occupations.
[89] This is equivalent to three to five days of software engineering time.
[90] [✂]; [✂]; [✂]; Name Withheld 22 response to the June 2025 Consultation, p.1; [✂].
[91] Francis, S. response to the June 2025 Consultation, pp.1, 3-5; Klaushofer, A. response to the June 2025 Consultation, p.1; Luethje, Y. response to the June 2025 Consultation, p.1; Mack, S. (Dr) response to the June 2025 Consultation, p.1; Name Withheld 22 response to the June 2025 Consultation, p.1; Name Withheld 26 response to the June 2025 Consultation, pp.1-2; [✂]; [✂]; Thompson, A. response to the June 2025 Consultation, p.1; Together Declaration Campaign response to the June 2025 Consultation, pp.1-2.

innovation if livestreaming providers are required to divert resources away from new functionalities in order to implement this measure.[92]

2.79 In our June 2025 Consultation, we considered the risk of market exit as a result of our measures on livestreaming for all users to be low. We maintain that the negative market impacts of this measure are minimal as the cost of adding the reporting category is relatively small. This will limit the impact on small services and on innovation in the market.

2.80 Industry stakeholders also highlighted that livestreaming providers will already have livestreaming safety tools which are tailored specifically to their service, at a lower cost than implementing our proposed reporting category.[93] The measures included in our Codes of Practice are recommendations for services to comply with their safety duties (and represent a safe harbour). They do not prevent services from using alternative safety tools which they can demonstrate to provide an equal level of effectiveness at mitigating the harms targeted through this measure.

Rights impact

Freedom of expression/Freedom of association

2.81 In considering this measure we have had careful regard to the impact it could have on users' rights to freedom of expression and association.

2.82 A number of stakeholders raised concerns that our proposals risk chilling speech, would cause users including journalists, researchers, and campaigners to self-censor, would reduce diversity and lead to incorrect takedown of important content.[94] [95] The Together Declaration Campaign raised concerns about the impact our measures could have if they led stakeholders to withhold information or discourage and delay reporting.[96] A number of industry stakeholders and individuals raised concerns that this reporting category could potentially be exploited to end a livestream for malicious reasons, which could have an impact on freedom of expression.[97]

2.83 We do not consider that this measure directly interferes with freedom of expression or association rights, and the indirect effects are likely to be minimal.[98]

---

[92] [✂]; [✂].
[93] [✂]; [✂].
[94] Francis, S. raised concerns that our proposals risk chilling speech and would cause users including journalists, researchers, and campaigners to self-censor, and would disproportionately affect marginalised voices, reduce diversity in public debate, and wrongly remove investigative reporting, whistleblowing and research that used sensitive images. Francis, S. response to the June 2025 Consultation, p.3.
[95] Stakeholders who also raised similar themes included: [✂] and Mack, S. (Dr), response to the June 2025 Consultation, p.2. We provide further details in the relevant chapter in annex 1: Further stakeholder responses.
[96] Together Declaration Campaign response to the June 2025 Consultation, p.1.
[97] Are, C. (Dr) response to the June 2025 Consultation, p.3; [✂]; [✂]; [✂]; [✂]; [✂]; [✂].
[98] In our June 2025 Consultation, we said a potential indirect restriction of rights arises where a service, as a result of report(s) labelled as "risk of imminent physical harm" (and possibly also based on other information reasonably available to the service provider such as meta data or negative feedback from users excludes the livestream from recommender feeds in accordance with our recommender systems measure. This is particularly significant where the reported livestream is not in fact illegal content, as removal from recommender feeds may have a negative impact on real-time engagement with the livestream. The impact would be as a result of the service providers decision and not a direct result of the measure. This interference would be temporary if the livestream were reinstated (for example, after review). Given the seriousness of potential harm, we do not consider the interference to be disproportionate.

2.84 We recognise that it is possible that some service providers may take an overly cautious approach when moderating content reported using this reporting category which could have an impact on a users' freedom of expression. However, providers are unlikely to take content down without moderating it because to do so would incentivise malicious reporting and impact on their users' amenity.

2.85 We have aimed to mitigate the risks of over moderation by giving clear guidance in the Illegal Content Judgements Guidance, which is intended to help services make judgements based on reasonably available information. There are also other measures in place to ensure there is a mechanism to appeal incorrect removal decisions.[99]

2.86 Where reported content is illegal (or illegal content proxy) and removed, this reflects a service's duty under the Act, and we do not consider this to be a disproportionate interference with freedom of expression.[100]

2.87 This could also have positive impacts on freedom of expression where users feel more confident engaging with livestreams if they know that content of this nature can be reported easily and that they are less likely to encounter it.

2.88 In our June 2025 Consultation, we said that given the limited direct interference posed by the measure, and the safeguards to mitigate the indirect interference referenced in paragraph 2.85, we consider that the potential interferences are proportionate having regard to the benefits of the measure in assisting service providers to take appropriate moderation action in relation to illegal content.

2.89 The measures are intended to improve the quality of information available to services to take appropriate action in relation to severely harmful content – nothing in our proposals requires service providers to take down lawful content.

Privacy

2.90 We note that the Act requires service providers to take appropriate action in response to relevant complaints, and that this duty was already implemented in the Illegal Content Codes on complaints handling. As such, the measure does not change the impacts on privacy.

2.91 Dealing with reports inevitably involves processing personal data, which may be human reviewed and may have an impact on users' rights to privacy and their rights under data protection law. The privacy of those depicted in livestreams, who may not be users, may also be affected.

2.92 In our June 2025 Consultation we said that we consider that users' privacy rights would not be disproportionately impacted by the measure, as service providers are required to comply with the relevant data protection legislation.[101] In moderating, providers would have to apply data protection principles, including (for example) those of accuracy and data

[99] Illegal Content Codes, Reporting and complaints, Measures ICU D1, D2, D8, D9, D10.
[100] Illegal content proxy for user-to-user services is 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 proxy for search services is search content that has been identified in the provider's publicly available statement for the service as being subject to appropriate moderation action, where the provider is satisfied that illegal content is included within that kind of content (including but not limited to priority illegal content).
[101] This includes the Data Protection Act 2018, the UK General Data Protection Regulations (UK GDPR) and, where relevant, the Privacy and Electronic Communications (EC Directive) Regulations 2003 (PECR).

minimisation. No evidence presented in consultation responses gives us grounds to change this assessment.

2.93 To the extent there is an interference with users' privacy, we consider this interference proportionate given the significant benefits of content moderation in reducing harm.

Who this measure applies to

2.94 In our June 2025 Consultation, we proposed that this measure should apply to all user-to-user services which feature one-to-many livestreaming as a primary or additional functionality and are at medium or high risk for one or more of the select list of harms. These harms include terrorism; grooming or image-based CSAM; assisting or encouraging suicide; hate; or harassment, stalking, and abuse offences. Following the consultation, we now consider that in addition to these harms, the measure should also apply to services that are at medium to high risk for content relating to animal cruelty. We outline the reasons for this is paragraphs 2.62-2.64. In addition, suicide and self-harm are now considered as one kind of illegal harm.[102]

2.95 We prioritise these specific illegal harms as, in our assessment, they are the most likely to need to be addressed quickly and in real time when broadcast on a livestream. Our Risk Assessment Guidance also identifies livestreaming as a key risk factor for these particular harms. We consider it proportionate that this measure should apply to medium and high risk services regardless of size.

2.96 We proposed including small services in the scope of this measure as there is a risk that illegal content is displaced from large services if small services do not have an appropriate reporting category. We consider the cost of this measure to be manageable for smaller services as the cost of triage and moderation will likely increase with service size and the volume of reports that a service is expected to receive. This is also a proxy for the potential for harm on a service so benefits are increasing with cost.

2.97 The Marie Collins Foundation said the reporting category should be recommended for all livestreaming services regardless of risk level.[103] We consider a risk-based approach to measure development to be appropriate in this case, given that certain services present a low risk for these select harms. As outlined in the Act, in this approach we account for risk levels, nature of harms and capacity of services. It would be disproportionate to include low risk services in the scope of the measure. The scope of this measure captures harms described in our Illegal Harms Register and we expect the majority of livestreaming services to be within scope.

Inclusion of animal cruelty

2.98 While the inclusion of animal cruelty may bring additional services into scope of the measure, it is likely that the majority of livestreaming services with risks of animal cruelty content would already have been in scope under our original proposal. This is because many of the factors which give rise to a risk of animal cruelty occurring on a service are also risk factors for the other harms in scope of this measure.[104]

[102] See chapter 1: Introduction, our duties and navigating the Statement for the reasoning.
[103] Marie Collins Foundation response to the June 2025 Consultation, p.9.
[104] Ofcom (2024), Illegal Harms Register of Risks.

2.99 We also received stakeholders requesting other harms be included in the scope of the measure. As we have not changed the scope beyond the inclusion of animal cruelty, we have included our response to the requests in annex 1: Further stakeholder responses.

Conclusion

2.100 Based on our reasoning set out in the previous sections, and having reviewed the June 2025 Consultation responses, we have decided to proceed with this measure for services that offer livestreaming as a functionality.

2.101 We expect this measure to be beneficial for user safety by enabling more effective user reporting of content that involves imminent physical harm and is therefore suspected illegal content. By allowing users to submit a report quickly without having to choose between harm categories, the measure reduces friction at the point of reporting. In turn, it also provides a clear signal to services that a report relates to imminent harm, enabling them to prioritise and take appropriate action more quickly in response.

2.102 We consider that the costs and impact on rights are relatively low when weighed against the benefits to users and that the measure is therefore proportionate.

2.103 We are applying this measure for all user-to-user services which offer one-to-many livestreaming services as a primary or additional functionality and that are medium or high risk for one of more of the following harms: terrorism; grooming or image-based CSAM; assisting or encouraging suicide and serious self-harm; hate; harassment, stalking, and abuse offences; or animal cruelty.

2.104 The full text of the measure and the services in scope can be found in our Illegal Content user-to-user Codes of Practice for Terrorism, CSEA and for other duties and is referred to as ICU D17.

Measure on human content moderation

2.105 In the June 2025 Consultation, we proposed a measure which set out that providers should ensure human moderators are available whenever users can livestream using the service (ICU C16).

2.106 We stated that this measure would apply to all user-to-user services offering one-to-many livestreaming at medium or high risk of one or more of the following harms: terrorism content, grooming or image-based CSAM, encouraging or assisting suicide, hate, animal cruelty, or harassment, stalking, threats, and abuse offences. This included services that only provide livestreaming and those where it is an added feature.

Our decision

2.107 After reviewing consultation responses, we have identified both uncertainty over the likely effectiveness of the proposed measure and concern that the costs outweigh the potential benefits to users. We have therefore decided not to proceed with this measure. We set out our reasoning for our decision in the following paragraphs.

Our reasoning

2.108 We received feedback both opposing and supporting the measure and have considered the stakeholder responses received.

Feedback requesting further clarification

2.109 Some stakeholders asked that we provide further detail on the number of human moderators that a service should deploy, relative to risk and size of the service.[105]

2.110 The Internet Watch Foundation (IWF), for example, said guidance should set out a proportionate number of human moderators for all in-scope services. It said this number must reflect the size and risk of the service and include moderators for a dedicated CSEA reporting branch.[106] South West Grid for Learning (SWGfL) made a similar point.[107]

2.111 The National Society for the Prevention of Cruelty to Children (NSPCC) said that we must share the minimum standards required for services when resourcing moderation teams on livestreaming services.[108]

2.112 Meta Platforms, Inc. (Meta) asked for clarification regarding human moderators being "available at all times that [livestreaming] is available to users", raising the following points:

- that human moderators may be on call but may not be in a position to moderate livestream content;
- that there was no clarity on the role of moderators, just that they are "available";
- that it would be disproportionate for all livestreamed content to be subject to human review and that we should clarify when human moderators are expected to moderate livestream content.[109]

2.113 The Molly Rose Foundation said that we should set clear expectations for what actions moderators should take.[110]

Feedback on performance targets and prioritisation

2.114 Stakeholders noted that the speed of human review was an important element to the effectiveness of the measure.[111] They said that we should set out expectations on performance targets as part of this measure.

Feedback on proactive technology

2.115 Some stakeholders pointed to the ongoing advances in the development of proactive technology and noted that technology for livestreaming has improved and is being adopted.[112]

---

[105] IWF response to the June 2025 Consultation, p.6; Meta response to the June 2025 Consultation, p.4; Molly Rose Foundation response to the June 2025 Consultation, p.13; NSPCC response to the June 2025 Consultation, p.4.
[106] IWF response to the June 2025 Consultation, p.6.
[107] SWGfL response to the June 2025 Consultation, p.8.
[108] NSPCC response to the June 2025 Consultation, p.4.
[109] Meta response to the June 2025 Consultation, p.4.
[110] Molly Rose Foundation response to the June 2025 Consultation, p.13.
[111] Antisemitism Policy Trust response to the June 2025 Consultation, p.1; Lucy Faithfull Foundation response to the June 2025 Consultation, p.2; Online Responsibility Network response to the June 2025 Consultation, p.2; SWGfL response to the June 2025 Consultation, p.8; [✂]; University College London (UCL) Gender and Tech Research Lab response to the June 2025 Consultation, p.4.
[112] 5Rights response to the June 2025 Consultation, p.7; CCO response to the June 2025 Consultation, pp.7, 8; Meta response to the June 2025 Consultation, p.3; Name Withheld 11 response to the June 2025 Consultation, p.1; NSPCC response to the June 2025 Consultation, p.3; ODDA response to the June 2025 Consultation, p.2; [✂]; UK Finance response to the June 2025 Consultation, p.3; [✂].

2.116 [✂], the Online Dating and Discovery Association (ODDA), [✂], and Meta disagreed that human moderation leads to better content moderation outcomes.[113] They argued that we should provide services with the flexibility to deploy both human and automated moderation of livestreams.

2.117 Stakeholders, including UK Finance, the Children's Commissioner's Office (CCO), and the NSPCC, also supported the use of proactive technology for livestreaming, noting it could be faster and more cost effective than human moderation.[114]

2.118 Name Withheld 11 said that proactive technology could reduce the cost for small services.[115]

Feedback on impact on small and medium services

2.119 Multiple respondents said that cost and resource demands associated with the proposed measure were challenging for smaller services.[116]

2.120 [✂]. Other stakeholders discussed barriers to market entry for new services lacking the necessary resources to implement content moderation systems for livestreams.[117]

2.121 One individual said that they disagreed with our view that most services with a livestreaming capacity are medium or large services which already have high fixed costs.[118]

2.122 BILETA acknowledged that while smaller services may lack all the necessary resources for moderation, they are "rightly included within scope" and that "Ofcom's recognition of this imbalance is welcome".[119]

Other feedback received

2.123 Some stakeholders acknowledged the value of having human moderators, in addition to automated systems, to assess and review harms on livestreams. Stakeholders noted the value of having human moderators.

2.124 They highlighted that human moderators could enable swift takedown of illegal content, are better able to identify complex harms such as grooming, coercion, distress, or manipulation, that proactive technology cannot. Feedback also noted that human moderators play a critical role in ensuring accuracy, [✂] and [✂].[120]

---

[113] Meta response to the June 2025 Consultation, pp.3, 5; ODDA response to the June 2025 Consultation, p.2; [✂]; [✂].
[114] CCO response to the June 2025 Consultation, pp.5, 6; NPSCC response to the June 2025 Consultation, p.4; UK Finance response to the June 2025 Consultation, p.5.
[115] Name Withheld 11 response to the June 2025 Consultation, p.1.
[116] BILETA response to the June 2025 Consultation, p.5; CCIA response to the June 2025 Consultation, pp.3, 4; Evangelical Alliance UK response to the June 2025 Consultation, p.2; Francis, S. response to the June 2025 Consultation, pp.3, 4; [✂]; [✂]; Klaushofer, A. response to the June 2025 Consultation, p.3; Name Withheld 11 response to the June 2025 Consultation, p.1; [✂]; Open Rights Group response to the June 2025 Consultation, pp.8, 9; [✂]; Together Declaration Campaign response to the June 2025 Consultation, p.1; [✂].
[117] [✂]; Open Rights Group response to the June 2025 Consultation, p.8.
[118] Klaushofer, A. response to the June 2025 Consultation, p.2.
[119] BILETA response to the June 2025 Consultation, p.4.
[120] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.6; BILETA response to the June 2025 Consultation, p.4; Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p.3; IWF response to the June 2025 Consultation, p6; [✂]; [✂]; [✂].

2.125 We received other feedback relating to the proposed measure.[121] As we are no longer proceeding with this measure, we want to thank the large number of stakeholders, including service providers and public sector bodies that have taken the time to engage with this proposal and supplement our evidence base.

Our response

2.126 Responses to our June 2025 Consultation raised important concerns about the proposed measure requiring services to ensure the availability of human moderators for livestreams. Stakeholders highlighted that a recommendation framed solely around availability would not, on its own, be meaningful without accompanying expectations about the speed of moderation or the proportion of content subject to human review.

2.127 We agree with stakeholders that the effectiveness of a content moderation function should be assessed primarily by the outcomes it delivers, rather than describing how that function is configured. In this context, we have concluded that it would not be appropriate to introduce prescriptive recommendations on the availability of human moderators for livestreams.

2.128 We recognise that, without detailed specification on moderation speed or coverage, services may find it difficult to determine whether a recommendation focused on moderator availability alone would meet regulatory expectations or deliver effective risk mitigation.

2.129 However, we are not currently able to provide detailed or prescriptive thresholds in these areas, given the significant variation in service models, scale, and functionality across the livestreaming sector.

2.130 The Illegal Content Codes already establish an outcomes-focused framework for content moderation. The content moderation measures (ICU C1, ICU C2) recommend that providers use effective systems and processes to review, assess and take down illegal content swiftly, including livestreams. Measure ICU C1 (Having a content moderation function) recognises that this may involve automated moderation, human moderation or a combination of both, and notes that human judgement may be needed where contextual assessment is necessary.

2.131 Measure ICU C2 (Swift take down of illegal content) further recommends that illegal content is taken down swiftly once identified. Together, measures ICU C1 and ICU C2 are deliberately not prescriptive about the tools used, recognising that effective moderation may, in some cases, recommend human involvement to support both accurate assessment and prompt enforcement.

2.132 In addition, measures ICU C3 and C4 recommend that large and/or multi-risk services set internal content policies and performance targets for their content moderation functions which balance the need for speed of decision making with the need for content moderation decisions to be accurate. These measures stipulate that services should resource and organise their moderation functions to give effect to these internal content policies and performance targets (ICU C6). Furthermore, ICU C5 is intended to ensure that large and multi-risk services take a structured approach to prioritising suspected illegal content, having regard to harm severity and user exposure.

[121] We outline some of this stakeholder feedback in the relevant chapter in annex 1: Further stakeholder responses.

2.133 When taken together, the existing measures described in paragraphs 2.130 and 2.132 should already go a long way to ensuring effective moderation of livestreamed content. We have concluded that the measure proposed in the consultation would not materially improve the existing framework. On the other hand, introducing uniform or prescriptive requirements on human moderators, moderation speed, or coverage risks imposing disproportionate burdens, constraining innovation, and not reflecting the diversity of livestreaming services.

2.134 A more prescriptive approach could also create unintended consequences: targets set too low may offer little protection, while targets set too high could place unnecessary burdens on services, potentially reducing user safety and limiting smaller providers' ability to operate.

2.135 Human moderators are still important, particularly where contextual judgement is needed, but effective moderation will typically require a proportionate combination of human and automated approaches as technologies continue to evolve.

2.136 In the absence of detailed guidance on moderation speed or human review levels, we currently consider the existing measures in the Illegal Content Codes to provide a sufficient and proportionate basis. This reflects the broader principles-based approach set out in our December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement).

Conclusion

2.137 Based on the June 2025 Consultation responses received and our analysis of this feedback, we have decided not to proceed with this measure.

2.138 As stated in paragraph 2.135, we still consider that human moderation should be an important element to service's livestreaming function, but relevant services will be able to decide how they moderate livestreams as part of their existing content moderation measure(s).

2.139 We will keep the moderation of livestreams under review to ensure services are making prompt judgements on content that is suspected to be illegal. We may reconsider our position in the future.

Measure to protect children

2.140 In the June 2025 Consultation, we proposed that user-to-user services implement restrictions when children in the UK undertake a one-to-many livestream broadcast. These included that users cannot post comments, send gifts to children or post reactions during children's livestreams; and that viewers should be prevented from capturing or recording videos or images of children's livestreams via in-service functionalities, and via third-party services where technically possible.

2.141 We stated that this measure would apply to all viewers of UK children's livestreams, regardless of whether the viewer is logged into an account on a service, and that services should use highly effective age assurance to apply these measures to users that have not been determined to be adults.

Our decision

2.142 Earlier this year, the UK Government announced its consultation on measures to prepare children for the future in an age of rapid technological change.[122] This included potential age restrictions on social media and other services, such as gaming sites and AI chatbots, and restrictions on harmful functions, including livestreaming.

2.143 Later, the UK Government announced its intention to restrict under-16s' access to social media. It also announced its intention to restrict functions such as livestreaming and stranger communication with children for under-16s. The UK Government stated that these restrictions will apply to a wider range of online services, including on gaming sites.[123]

2.144 Due to this announcement, we have decided not to proceed with this measure at present due to the similarity of the proposals.

2.145 We do not consider that it would be appropriate for us to include this measure in our Illegal Content Codes as it currently stands, given that services may be required to make significant further changes following potential new legislation.

2.146 We thank all stakeholders who submitted responses; we will take this feedback into account as part of any future proposals or decisions on livestreaming. Published consultation responses can be found on Ofcom's website.[124]

[122] UK Government (2026), Growing up in the online world: a national consultation - GOV.UK.
[123] UK Government (2026), Social media to be banned for under-16s in landmark government move to give kids their childhood back - GOV.UK, UK Government, 2026. Growing up in the online world: government response (July 2026).
[124] Ofcom (2025), June 2025 Additional Safety Measures Consultation.

3. User Sanctions

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

What is this chapter about?

This chapter sets out measures to ensure online services have effective and proportionate policies for sanctioning users who share, receive or view illegal content or content harmful to children. These sanctions include giving users warnings, suspensions or bans from a service, or restricting use of a service. The aim is to prevent repeat offending and reduce the future spread of harmful content online.

What decisions have we made?

Banning users for CSEA

We have decided that providers of user-to-user services should ban users who generate, upload, share or receive and view child sexual exploitation and abuse (CSEA) content, including child sexual abuse material (CSAM and proxy content). Providers should prevent banned users from returning to a service, but we are not prescriptive about the technical methods used to do so.

For adult users, the default expectation is a permanent ban, except in extremely limited and exceptional circumstances.

Providers should treat children that share CSAM in a more nuanced way. Providers may decide not to ban children who shared sexual images of themselves because they were groomed, or exchanged images of themselves consensually within an age-appropriate relationship. In other cases involving child users, providers should impose bans that are proportionate to the circumstances.

User sanctions policy

We have decided that providers should prepare and apply a sanctions policy in respect of UK users who generate, upload, or share illegal content or content harmful to children which is prohibited on the service. We are not prescriptive about the sanctions they apply, and how they apply them, as long as they have regard to the factors set out in our measures: the severity of potential harm, whether the user has previously generated, shared or uploaded relevant content, and the potential impact of the sanction on the user.

We have also made a number of other amendments to existing code measures to complement these user sanctions measures.

Having considered our cost estimates and potential benefits, we have decided that these measures should no longer apply to providers of small, low risk services, or single risk providers.

Our final Code measures are summarised in the table below:

Number in Codes	Recommended measure	Who should implement this
ICU H3	Banning users for CSEA	All user-to-user service providers.

Number in Codes	Recommended measure	Who should implement this
ICU H2	User sanctions policy for illegal harms.	Providers of user-to-user services that are large and/or multi-risk for illegal harms.
PCU H2	User sanctions policy for content harmful to children.	Providers of user-to-user services likely to be accessed by children that are large and/or multi-risk for content harmful to children and prohibit one or more specific kinds of PPC, PC and/or NDC.
ICU C4	Set and record performance targets (for its content moderation function)	Providers of large and/or multi-risk user-to user-services where take down is not technically feasible.
PCU C4		Providers of large and/or multi-risk user-to-user services which are likely to be accessed by children, where content and/or access level action is not technically feasible.
ICU D8	Set performance targets and resource itself to give effect to those targets for relevant complaints which are content appeals.	Providers of user-to-user services that are large and/or multi-risk for illegal harms.
PCU D8		Providers of user-to-user services likely to be accessed by children that are large and/or multi-risk for content harmful to children and prohibit one or more specific kinds of PPC, PC and/or NDC.
ICU D10	Action following determination for relevant complaints which are content appeals.	All user-to-user service providers.
ICU D10A	If a provider does not reverse a decision that content was illegal content or illegal content proxy following appeal (other than CSEA content or CSEA content proxy) but upholds, in whole or in part, a content appeal relating to the user being sanctioned, the provider should reverse the action taken against the user, and apply the appropriate sanction having regard to its sanctions policy.	Providers of user-to-user services that are large and/or multi-risk for illegal harms.

Number in Codes	Recommended measure	Who should implement this
PCU D10A	If a provider does not reverse a decision that content was content harmful to children or content harmful to children proxy but upholds, in whole or in part, a content appeal relating to the user being sanctioned, the provider should reverse the action taken against the user and apply the appropriate sanction having regard to its sanctions policy.	Providers of user-to-user services likely to be accessed by children that are large and/or multi-risk for content harmful to children and prohibit one or more specific kinds of PPC, PC and/or NDC.
ICU G1	Include all information mandated by the Act in terms and statements including its policy on banning users in relation to CSEA content or CSEA content proxy.	All user-to-user service providers.
ICU G1A	The provider should include in the terms of service provisions giving information about its policy on sanctions, addressing, at least, the sanctions which the provider may impose on users who generate, upload or share content that it determines to be illegal content and/or illegal content proxy, which is not CSEA content or CSEA content proxy.	Providers of user-to-user services that are large and/or multi-risk for illegal harms.
PCU G1A	The provider should include in the terms of service giving information about its policy on sanctions, addressing, at least, the sanctions which the provider may impose on users who generate, upload or share content that it determines to be relevant content that is harmful to children and/or content that is harmful to children proxy.	Providers of user-to-user services likely to be accessed by children that are large and/or multi-risk for content harmful to children and prohibit one or more specific kinds of PPC, PC and/or NDC.

Why have we made these decisions?

Banning users for CSEA

CSEA is a severe harm, and we consider that permanently removing adult offenders from services is a necessary and proportionate measure to mitigate this. The approach provides the strongest deterrent to future offending and offers the greatest protection to child users. A more nuanced approach to banning child users ensures that children are not unduly punished for sexual images they shared when they were groomed or for images shared within consenting age-appropriate relationships with another child.

User sanctions policy

Once content is identified as illegal or harmful to children through content moderation processes, there is an ongoing risk to users if no action is taken against the users who disseminated it. Without sanctions, there is also little deterrent to stop users from continuing to generate, upload and share this content.

These measures are therefore intended to prevent repeat violations and reduce the spread of harmful content by educating users, discouraging harmful behaviour and, where providers consider this appropriate, removing user access altogether.

Introduction

3.1 User sanctions are actions that a service provider may take against a user who has generated, uploaded, or shared illegal content or content harmful to children. These actions may include giving the user a warning, suspending or banning the user from accessing the service, or in any other way restricting the user's ability to use the service. The Online Safety Act (the Act) expressly states that the safety duties apply, where proportionate, to policies on user access, including blocking users from accessing the service.[125]

3.2 In the June 2025 Additional Safety Measures Consultation (June 2025 Consultation), we consulted on three new measures:

a) One measure recommending that providers should ban users who generate, upload, share, or receive child sexual exploitation and abuse (CSEA) content and take reasonable steps to prevent the user from returning to the service for the duration of the ban (ICU H3).
b) Two measures recommending that providers should prepare and apply a user sanctions policy:
    i) in respect of users who generate, upload, or share illegal content (ICU H2); and
    ii) in respect of users who generate, upload, or share content harmful to children that is prohibited on the service (PCU H2).

3.3 We also proposed related amendments to existing measures. As such, this chapter also covers amendments to the:

a) reporting and complaints measures on appropriate action – determination and following determination for content appeals (ICU D8 and D9, PCU D8 and D9, ICU D10, and PCU D10, including additions of measures ICU D10A and PCU D10A);

[125] Sections 10(4)(d) and 12(8)(d) of the Act. Section 8(3) provides that these duties extend only to the design, operation and use of the service in the UK.

b) terms of service measures on substance (ICU G1, including additions of measures ICU G1A and PCU G1A); and
c) content moderation measures on performance targets (ICU C4 and PCU C4).

3.4 This chapter is structured as follows:

a) The first section sets out our decision on our proposed measure on banning users for CSEA content;
b) The second section sets out our decision on our proposals that providers should have and apply a user sanctions policy; and
c) The final section sets our decisions on amendments to other Illegal Content or Protection of Children Codes measures to reflect the user sanctions measures.

3.5 In this chapter, we address the feedback we received and confirm our final decisions.

Measure on banning users for CSEA

3.6 In our June 2025 Consultation, we consulted on a proposal to recommend that service providers ban, and prevent from returning, users who have generated, uploaded or shared child sexual exploitation and abuse (CSEA) content and CSEA content proxy, and users who have received child sexual abuse material (CSAM) and CSAM proxy.[126]

3.7 We explained the aim of this measure was to remove CSEA perpetrators from services and reduce the risk of harm to children in the UK. We defined a 'ban' as a user being prevented from accessing the service, and any other regulated user-to-user services operated by the provider where it has reasonable grounds to believe that the user has access to those services. We proposed that bans should be permanent, apart from in limited exceptional circumstances where a provider considers a time-limited ban would be more appropriate.

3.8 We also proposed a range of options for how to deal with the issue of child users to ensure that children who perpetrate harm are appropriately sanctioned, while children who have been a victim of an offence or have shared images consensually within an age-appropriate relationship are not unduly punished.

3.9 We proposed that this measure be applied to all user-to-user services to offer the greatest protection to victims and survivors and to avoid the creation of some services where perpetrators are more able to cause harm.

Our decision

3.10 We have decided to broadly confirm the measure as proposed in the June 2025 Consultation. Based on stakeholder feedback, we have made some amendments to the measure. This includes amending the wording of the measure to permit providers not to ban users who report and delete CSAM after receipt. Further detail on these amendments

[126] All references to child sexual exploitation and abuse (CSEA) and child sexual abuse material (CSAM) in this chapter are to illegal CSEA content (which includes both grooming and CSAM content) and CSEA content proxy together, unless otherwise stated. For user-to-user services, content proxy is 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 of service prohibit the types of illegal content in question (here, this refers to illegal CSEA content and illegal CSAM content).

is set out in the section 'Amendments to the Codes to reflect user sanctions policy measures and user banning measure.'[127]

3.11 We also received significant stakeholder feedback regarding how children are treated under this measure. We have considered stakeholder feedback carefully and agree that the treatment of child users under this measure needs a nuanced approach.

3.12 As proposed in the June 2025 Consultation, we are confirming that providers may decide not to ban a child where they have reasonable grounds to believe that either of the following extenuating circumstances exist: the child shared self-generated indecent images (SGII) as a result of being groomed, or the child consensually sent SGII or received SGII consensually as part of an age-appropriate relationship. We have also decided that, where a provider has reasonable grounds to believe a user is a child, but one of these extenuating circumstances does not apply, the default position should not be a permanent ban as it is for adults. Instead, providers should reasonably determine an appropriate duration for the ban, ensuring the duration considered achieves the specific outcomes which we set out in paragraph 3.71.[128] Where a provider does not have reasonable grounds to believe that the user is a child, they should be treated as an adult under this measure.

3.13 Earlier this year, the UK Government announced plans to introduce new legislation to restrict under-16s' access to social media, restrict certain features on a wider range of services, and restrict those features by default for 16- and 17-year-olds. In taking the above decisions relating to this measure, we have considered the implications of this announcement and the continued necessity and proportionality of this measure. We remain of the view that the measure still presents significant benefits both for adults and children, as set out in paragraph 3.38.

3.14 We have also confirmed our proposed amendments to existing appeals and terms of service measures to support this banning measure. See section titled 'Amendments to the Codes to reflect User Banning measure' for a detailed explanation of these changes.

3.15 This measure is part of our Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes) for CSEA and is referred to as ICU H3. Amendments to measures ICU D10.3, previously ICU 10.4, and ICU G1.2, are explained in paragraphs 3.321-3.330.

How this measure works

Who should be banned

3.16 This measure is intended to ban users who carry out grooming, and who share and receive CSAM. It therefore recommends that users who generate, share, or upload CSEA content on the service should be banned from the service and prevented from returning.[129] Users who receive and view CSAM without promptly reporting and deleting it should also be banned from the service and prevented from returning. We confirm the definition of a ban remains the same as in our June 2025 consultation, as set out in paragraph 3.7.

[127] Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes), Measure ICU H3.4-H3.5.
[128] Providers may consider that a permanent ban is appropriate.
[129] Consistent with paragraph 11 of Schedule 4 to the Act, to follow this measure providers would need to take these actions in relation to the design and operation of the service in the UK, and as it affects UK users.

3.17 The measure applies to all forms of CSEA including CSAM and grooming offences.[130] We set out the rationale for this in paragraphs 3.48-3.49. We recognise that providers may choose to identify and take down CSEA content in accordance with their terms of service, rather than by making an illegal content judgement. As such, the measure also recommends that providers should ban a user where they identify that user to have posted CSEA content proxy.

3.18 We have amended the wording of the measure to clarify that a ban only applies to content that has been uploaded 'on the service'. This is to ensure that users who upload content to a service's reporting system are not banned, and that this measure therefore does not inadvertently deter users from reporting CSEA content. We confirm that reporting content to the service provider does not constitute generating, sharing or uploading that content on the service.

3.19 Following stakeholder feedback on the potential impact of banning users who receive CSAM without soliciting it (see paragraph 3.134), we have amended the measure so that providers may decide not to ban a user who has received and viewed CSAM but has promptly reported and deleted it, where deleting the content is possible. The measure specifies that, to avoid being banned, the user must have reported the content to the provider, an appropriate law enforcement agency, or any other dedicated CSAM reporting agency, such as the Internet Watch Foundation. We have decided providers should have discretion to ban users even where they have promptly reported and deleted the content, as we recognise that there may be occasions where a ban may be appropriate even in those circumstances. For example, where the provider has evidence the CSAM was solicited.

3.20 We recognise that providers may not have evidence of a user having promptly reported and deleted CSAM content at the time of applying a ban. We expect providers' appeals processes to allow users to provide evidence of such reporting, in which case providers may choose to overturn a ban if they are satisfied the user promptly reported and deleted the content in question. For more details on our appeals processes, see paragraphs 3.32-3.36.

3.21 Given that offences often take place across multiple services, we expect service providers to use all relevant available information when determining if a user received and viewed CSAM content, without appropriately reporting and deleting. We expect this to include any evidence as to whether the user opened or interacted with the content in any other way.

3.22 Providers should include clear and accessible information in the terms of service about the policy on banning users for CSEA, including the circumstances in which a user will be banned.[131]

Technical implementation of banning and preventing return

3.23 Service providers should take reasonable steps to prevent banned users from returning to the service. Providers have discretion over the technical means for preventing a user from returning.

---

[130] Ofcom (2024), Illegal Content Judgements Guidance (ICJG). See Child sexual exploitation and abuse (CSEA): Offences relating to child sexual abuse material (CSAM) (pp.73-86) and Child sexual exploitation and abuse (CSEA): Grooming and exploitation of children (pp.86-101).

[131] Illegal Content Codes, Measure ICU G1.2(d).

3.24 The measure applies to all user-to-user service providers, including those with unregistered users. See paragraphs 3.50-3.62 for further detail on the benefits and effectiveness of our approach).

Duration of ban – adult users

3.25 For adult users, we have decided to confirm Option 3 from the consultation – that users should be banned permanently as a default. This means that providers should by default ban permanently any user that the provider does not have reasonable grounds to believe is a child.[132] We explain how the measure works where the provider does have reasonable grounds to believe the user is a child in the following section: Our approach to children.

3.26 Providers may ban adults for a lesser duration in exceptional circumstances, where they consider a permanent ban not to be appropriate. In these cases, the ban should be of such duration as the provider reasonably considers appropriate, ensuring the duration of the ban reflects the nature and severity of the harm and protects UK users, particularly children, from harm by deterring the banned user, and other users on the service, from sharing CSEA content.

3.27 We do not specify the exceptional circumstances where a permanent ban may not be appropriate, as making these factors public could be utilised by perpetrators to avoid a ban. Our expectation is that there will be very few exceptional circumstances where a permanent ban would not be appropriate for an adult user. An example may be where the provider is confident that the user has no knowledge or culpability, such as where there has been a security compromise on the account, or where images have been shared as part of an age-appropriate, consensual relationship, for example, between a 17- and 18-year-old.

Our approach to children

3.28 In our June 2025 Consultation, we proposed three options to manage banning children from a service. We have decided to confirm Option B – ban children who commit CSEA harm unless in exceptional circumstances, though we now refer to these as extenuating circumstances for the reasons set out in paragraph 3.87.

3.29 Accordingly, providers may decide not to ban a user who they have reasonable grounds to believe is a child where one of the following extenuating circumstances exists: that the user has shared CSAM as a result of grooming, or they consensually shared SGII or received SGII consensually as part of an age-appropriate relationship.

3.30 Stakeholder feedback highlighted the nuanced approach needed to children. We agree that there are circumstances, outside of the two extenuating circumstances, in which a child may share or receive CSAM online without intentionally or knowingly causing harm. For this reason, we have decided to take a different approach to the duration of a ban for children. We are recommending that, where providers are banning a user that they have reasonable grounds to believe is a child, the default position should not be that the ban is permanent but rather providers should impose a ban of a duration that they reasonably consider to be appropriate in the circumstances. Providers should ensure the duration of the ban reflects the nature and severity of the harm, takes into account whether the user has previously shared CSEA content and whether they were previously banned for doing so, takes account of the potential impact of the duration of the ban on the user, and protects users,

[132] In other words, if the provider does not have reasonable grounds to believe the user is a child, they should be treated as an adult.

particularly children, from harm by deterring the user and other users from sharing CSEA content.

3.31 We nevertheless recognise that children can intentionally harm other children online, as set out in paragraph 3.92. As such, providers may therefore decide it is appropriate to apply a permanent ban to children in some cases. We provide a full analysis of our approach to children from paragraphs 3.72-3.109.

Appeals

3.32 Providers have a duty under section 21 of the Act to take appropriate action in response to complaints by users who have been suspended or banned or otherwise had their use of the service restricted because of posting illegal content.[133]

3.33 Under our current reporting and complaints measure (ICU D10), if a service provider determines on appeal that the relevant content is not illegal, the provider should reinstate their account and relevant associated content.

3.34 We have decided to amend the reporting and complaints measure to provide that an appropriate action providers can take on appeal is to overturn a ban under our measure, where the content was correctly identified as CSEA, but the appeal shows that one of the identified extenuating circumstances for children exists. We have also amended the measure to allow providers to determine the appropriate duration of a ban on appeal if they receive information which gives them reasonable grounds to believe the user is a child. This is outlined in our section titled 'Our approach to children' and in the section titled 'Amendments to the Codes'.

3.35 We have also decided to amend the reporting and complaints measures to permit providers to impose a different duration of a ban for adult users where the content was correctly identified as CSEA, but an exceptional circumstance exists.[134]

3.36 Appeals are an important safeguard for users' rights, and providers should communicate a ban and advise users of their right to appeal. However, we are choosing not to specify the approach providers should take to sharing this information because this will vary from service to service and case to case.

Our reasoning

Effectiveness and benefits

3.37 This measure seeks to protect users from the harm caused by CSEA. As set out in our Illegal Harms Register of Risks (Illegal Harms Register), CSEA encapsulates a number of online offences including causing or inciting a child to engage in sexual activity, sexual

[133] See the reporting and complaints measures (ICU D) in the Illegal Content Codes, which recommend how providers should enable UK users and affected persons (as defined in section 20 of the Act) to appeal, including against a user being banned. Most service providers will operate a global appeals mechanism, and it is for a user's country of location to determine how they are treated for the purpose of appeals.
[134] See measures ICU D10.3(c) and ICU H3.8. We have amended these measures to better reflect our policy position that a ban of a lesser duration should only be applied where an exceptional circumstance exists, and that it is a provider's choice whether to impose a ban of a lesser duration.

communication with a child and taking, making, distributing and possessing indecent photographs of a child.[135] Research shows the prevalence of this harm online.[136]

3.38 By banning those found to have committed a CSEA offence, we seek to reduce the risk of harm by:

- Removing a perpetrator's access to a service and their ability to identify, contact and abuse children, thereby disrupting their opportunities to perpetrate.
- Reducing the risk of escalating perpetrating behaviours associated with accidental exposure to CSAM and perpetrator to perpetrator networking, acting as a deterrent and providing a consequence for harmful actions.
- Reducing the prevalence of CSEA content on services by removing those known to share this material and reducing their ability to share it online.
- Reducing the risk of the wider public being exposed to CSEA content.

3.39 In our June 2025 Consultation, we set out our evidence for how banning users will be an effective way of achieving these policy aims. We explained that while early perpetrating can often be accidental, over time offending behaviours become entrenched and normalised. This results in persistent, conscious offending that can become compulsive or addictive. A 2021 study by Meta Platforms Inc. (Meta) of over 200 offending accounts showed "patterns of persistent, conscious engagement with CSAM and other minor-sexualising content".[137] Similarly, research shows that in offender samples, online CSAM offenders show increased risk of compulsive or addictive behaviours when compared to those who commit offences offline.[138]

3.40 The normalisation of abuse can result in an escalation in the severity of offences and an increased risk of harm to children with many offenders reporting first accessing CSAM accidentally or without having searched for it.[139] Over time, this initial access can be seen to escalate with perpetrators being seen to engage in increasingly harmful sexual activity including the abuse of other children and/or animals as the grooming progresses. Evidence suggests viewing CSAM can increase the risk a perpetrator will contact a child.[140]

3.41 CSEA can have a significant impact on victims and survivors. This can include long term psychological, emotional, and physical effects, difficulties in forming and maintaining relationships and feelings of guilt, shame, and self-blame that can persist into adulthood. In the Protect Children Global Survivor Survey 2024, 84% of respondents reported long term consequences stemming from the abuse that they had suffered including depression (55%),

---

[135] Ofcom (2024), Illegal Harms Register of Risks (Illegal Harms Register), p.88.
[136] The IWF 2025 Annual Report showed a 6% increase year on year on reports assessed and a 7% year on year increase in the number of reports confirmed as being CSAM. IWF, 2025. Annual Report. Similarly, the 2025 Childlight 'Into the Light' index, it was estimated that 19.6% of children will experience online solicitation – including unwanted sexual contact, grooming, or sexual harassment – before the age of 18. Childlight, 2025. 2025 Into the Light Index on Global Child Sexual Exploitation and Abuse: Supplemental Thematic Analysis Report.
[137] A 2021 Meta study into 200 accounts showed perpetrators repeated and were persistent in their abuse of children. Meta (Buckley, J., Andrus, M., Williams, C.), 2021. Understanding the intentions of Child Sexual Abuse Material (CSAM) sharers.
[138] ECPAT International, 2018. Trends in online child sexual abuse material.
[139] Protect Children, 2026. CSAM perpetrator Research Project: Findings from a Survey of CSAM Perpetrators on Digital Platform Use and Design (Tell Me More About Tech).
[140] Ofcom (2026), Illegal Harms Register of Risks, p.58.

anxiety disorders and panic attacks (54%), and physical symptoms such as chronic pain and insomnia (38%).[141]

3.42 Based on the evidence outlined, we said that offenders who remain on a service continue to present a severe risk of harm to children. We therefore suggested that banning perpetrators would be a proportionate and effective way of preventing offending and reducing the risk of behaviours escalating and causing further harm. On this basis and given the severe harm CSEA causes, prevalence of CSEA online, and the efficacy of this measure in combatting CSEA, we argued that this measure would deliver very significant benefits. Nonetheless, we recognised the potential limitations of the measure, in terms of perpetrators seeking to circumvent a ban and the potential difficulties of preventing return to the service.

3.43 The evidence presented in response to the June 2025 Consultation responses broadly validates our assessment of the benefits of this measure. No evidence was presented by stakeholders which fundamentally undermines our assessment that the measure will be effective and beneficial. We received broad support for the measure from industry, civil society and government stakeholders such as the Marie Collins Foundation, the Canadian Centre for Child Protection (C3P), the Internet Watch Foundation (IWF), [✂], and [✂].[142] Some industry stakeholders also highlighted that our proposals largely reflect some existing industry practice.[143]

3.44 While stakeholder responses have not led us to substantially revise our assessment of the benefits of this measure, several stakeholders provided feedback or requested clarity on specific points of detail which could affect its efficacy. We discuss these points in the following sections.

3.45 We have considered how the UK Government's announcement of restrictions on social media use and certain features for under 16s, alongside its other proposals, may affect the proposed benefits of this measure. We recognise that the introduction of these restrictions will mean that fewer children should be able to access social media services in scope of this measure and that this will reduce the ability of offenders to cause harm to children. Nonetheless, we are of the view that the measure continues to present significant benefits in protecting UK users from harm.

---

[141] Protect Children, 2024. Our Voice: Experiences of Adult Survivors of Child Sexual Abuse and Exploitation Across 29 Languages.

[142] Canadian Centre for Child Protection (C3P) response to the June 2025 Additional Safety Measures Consultation, p.15; Check My Ads Institute response to the June 2025 Additional Safety Measures Consultation, p.10; Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Additional Safety Measures Consultation, p.7; Department of Justice for Northern Ireland response to the June 2025 Additional Safety Measures Consultation, p.13; Digital Resilience in Education (Welsh Government) response to the June 2025 Additional Safety Measures Consultation, p.12; [✂]; [✂]; Internet Matters response to the June 2025 Additional Safety Measures Consultation, p.12; Internet Watch Foundation (IWF) response to the June 2025 Additional Safety Measures Consultation, p.43; Marie Collins Foundation response to the June 2025 Additional Safety Measures Consultation, p.9; [✂]; National Center for Missing & Exploited Children (NCMEC) response to the June 2025 Additional Safety Measures Consultation, p.6; [✂]; [✂]; Scottish Government response to the June 2025 Additional Safety Measures Consultation, p.3; Social Media Animal Cruelty Coalition (SMACC) response to the June 2025 Additional Safety Measures Consultation, p.13. The Commissioner Designate for Victims of Crime for Northern Ireland said that the measure was likely to be effective in creating friction and modifying user behaviour.

[143] [✂]; [✂]; [✂].

3.46 This is because the measure will restrict the ability of adult offenders to share CSAM with other adult users, reducing the circulation of CSAM between adult users online and address the harm that is caused to the victims and survivors in these images, even where the victims and survivors themselves may not be able to see the images.

3.47 In relation to those aspects of the measure that relate to child users, these continue to be an important part of the overall effectiveness of the measure for the following reasons:

- The Government's proposed new legislation does not impact the importance of this element of our measure, as the spread of CSAM may take place between children over the age of 16, and between adults. Under current legislative proposals, children aged 16 and 17 will still be able to access social media and therefore remain at risk from the harms caused by CSEA. Evidence attests to the continued risk of harm to this age group; for example, research by Snap Inc. showed that 41% of children aged 16 and 17 are reported to have lost control of their intimate images to someone else.[144] This measure will act as an important protection for these children from the harm caused by CSEA.
- Under the new legislation, it may still be possible for children under 16 to access other services, such as direct messaging services, where they may be at risk of online CSEA. This measure applies to all user-to-user services and therefore offers protection to children of all ages on those services which will still be accessible by children under 16.

Applying the measure to all forms of CSEA

3.48 In our June 2025 Consultation, we proposed to apply this measure to all types of CSEA because:

- All types of CSEA can have a significant negative impact on victims and survivors. Applying the measure to all types of CSEA therefore offers children the greatest level of protection from harm.
- Perpetrators often commit multiple CSEA harms against a victim and offences are often interconnected.[145] For example, grooming is often used to coerce victims to create SGII or to engage in online or offline sexual activity.[146]
- To only ban for certain CSEA-related harms may incentivise perpetrators to shift their perpetrating to circumvent a ban. For example, a ban that only applies to known CSAM may encourage perpetrators to seek novel CSAM from victims.[147]

3.49 Civil society stakeholders agreed with our proposal for the measure to apply to all types of CSEA.[148] The Marie Collins Foundation said it agreed with applying the measure to those who receive CSAM, as well as those who send it.[149] In light of these responses, we remain of the view that the measure should apply to all types of CSEA.

---

[144] Snap Inc., 2023. Digital Wellbeing Index.
[145] Sections 4 and 5 of Ofcom's Illegal Content Judgements Guidance set out these offences.
[146] Ofcom (2024), Illegal Harms Register of Risks, p.66.
[147] We use known CSAM to refer to CSAM that has previously been identified and is known to law enforcement/hash lists.
[148] IWF response to the June 2025 Consultation, p.44; We Are Survivors response to the June 2025 Additional Safety Measures Consultation, p.7. IWF noted the benefits of this approach because of the impact that all forms of CSEA have on victims, the connected nature of CSEA offences and the need to prevent perpetrators finding loopholes.
[149] Marie Collins Foundation response to the June 2025 Consultation, p.9.

Technical implementation of banning and preventing return

Preventing return

3.50 In our June 2025 Consultation, we explained why preventing return is an important part of the measure and outlined our reasons for proposing that providers should have discretion over the technical means of banning and preventing return. Several civil society and industry stakeholders, including Barnardo's and Google LLC., welcomed the discretion given to providers.[150]

3.51 Stakeholders from both industry and civil society – including Google LLC., [✂], the IWF, and the National Society for the Prevention of Cruelty to Children (NSPCC) – sought clarity on the "reasonable steps" we expect providers to take to prevent users from returning to the service.[151]

3.52 Having considered the responses we received, we are not changing our position from the June 2025 Consultation. We do not consider it appropriate to specify the reasonable steps providers should take to prevent return, nor do we have the evidence to do so. This is due to both the range of services to which the measure applies and the ever-changing nature of technology in this area. We emphasise that in most cases there will be a range of steps a provider could take that may be considered reasonable, and that providers should determine what is reasonable for their own service considering several factors, such as the size of the service, its risk level, and whether users are unregistered. We will consider whether the steps a provider has taken are reasonable on a case-by-case basis.

3.53 We have therefore decided to confirm that providers should take reasonable steps to prevent users from returning, and that providers will be able to choose what those steps are.

3.54 We consider that giving providers discretion over the methods to prevent return increases the effectiveness of the measure because:

- There are a variety of different methods providers may use to prevent return, with best practice likely to include multiple different technologies or indicators. We note that different methods vary in costs and effectiveness.
- There is unlikely to be a one-size-fits-all approach as the measure applies to all user-to-user services. Giving flexibility allows providers to implement the measure in a way that is most effective for the size and risk level of their service.
- The technology available to achieve the outcome of the measure is ever-changing. We therefore expect the methods providers use to change as new and improved technologies emerge.
- Being prescriptive on how service providers implement this measure could inadvertently alert perpetrators how to circumvent a ban and reduce its effectiveness.

---

[150] Barnardo's response to the June 2025 Additional Safety Measures Consultation, p.14; Google LLC. response to the June 2025 Additional Safety Measures Consultation, p.28; [✂]; Online Dating and Discovery Association (ODDA) response to the June 2025 Additional Safety Measures Consultation, p.7.

[151] Google LLC. response to the June 2025 Consultation, p.30; IWF response to the June 2025 Consultation, p.44; [✂]; [✂]; National Society for the Prevention of Cruelty to Children (NSPCC) response to the June 2025 Additional Safety Measures Consultation, p.39.

Applying the ban across all services operated by a provider

3.55 We proposed that service providers also ban users from all other services operated by the provider which it has reasonable grounds to believe the user has access to. Some civil society stakeholders, including C3P, the IWF and the NSPCC, agreed with our proposals to apply a ban across all services operated by a provider.[152] One provider said that they already take a multi-layered approach to preventing banned users from returning, combining enforcement action, detection technology, account management and work with external partners.[153]

3.56 Some stakeholders raised concerns about applying a ban across other services. Google LLC. noted the impact this would have in a user's day to day life, for example if they were denied access to additional or essential services outside the scope of the Act.[154] A stakeholder highlighted technical challenges in applying bans across distinct systems or platforms, and potential impacts on services and users outside the UK.[155]

3.57 The measure as set out in the draft Illegal Content Codes did not use the defined term "service" in the definition of a "ban".[156] We have made clear in the final measure that references to "service" are to the defined term, which means a regulated user-to-user service.[157] This should address stakeholder concerns of a user being banned from services outside of the scope of the Act,[158] or impacting services and users outside of the UK.[159]

3.58 As set out within the Illegal Harms Register, offenders can 'platform-hop' between services to exploit different functionalities, enabling them to identify and contact children, engage in grooming conversations and move interactions with children to more private spaces to avoid detection. Defining "banning" to capture other services operated by the provider therefore helps to address cross-platform risks and offer greater protection to children. Providers have discretion over the technical means they use to implement this aspect of the measure, and they should only ban a user from other regulated user-to-user services they operate where they have reasonable grounds to believe a user has access to them.

---

[152] A member of the Philippine Survivor Network response to the June 2025 Additional Safety Measures Consultation, pp.19-20C3P response to the June 2025 Consultation, p.15; IWF response to the June 2025 Consultation, p.44; NSPCC response to the June 2025 Consultation, p.39. NSPCC said that it welcomes this aspect of the measure as it said many of the larger platforms are owned by the same company.

[153] [✂].

[154] Google LLC. response to the June 2025 Consultation, p.31.

[155] [✂].

[156] The word "service" appeared unformatted, whereas defined terms in the Illegal Content Codes are either emboldened, or emboldened and italicised.

[157] Section 3(1) of the Act defines a user-to-user service. Section 4(2) of the Act defines a regulated user-to-user service as a user-to-user service that has links with the UK and is not exempted as provided for by Schedules 1 or 2. Sections 4(5) and 4(6) of the Act set out the circumstances in which a service will have links with the UK.

[158] See Schedule 1 to the Act.

[159] Though the measure applies to all users, it applies only to the design and operation of the service in the UK, or as it affects UK users of the service (paragraph 11 of Schedule 4 to the Act). Providers' illegal content safety duties also extend only to the design, operation, and use of the service in the UK (section 8(3) of the Act).

Risk of circumvention

3.59 Some civil society stakeholders raised concerns around the risk of circumvention, for example that users may create new accounts to rejoin the service.[160] The British and Irish Law Education and Technology Association (BILETA) raised concerns about the use of virtual private networks (VPNs) to circumvent a ban.[161]

3.60 We recognise the potential limitations of the measure, in terms of perpetrators seeking to circumvent a ban and the potential difficulties of preventing return to the service. Nonetheless, we consider the measure still to be effective in that it increases friction to the point that perpetrators are unable to, or are disincentivised from, trying to return to the service.

Unregistered users

3.61 We recognise that there are difficulties with applying this measure in some cases, including providers which operate multiple services where it may not be clear which services a user has access to, services which operate in a decentralised manner, or services which allow unregistered users. Middle Tech Coalition acknowledged the difficulties in banning unregistered users.[162]

3.62 We considered whether to exclude unregistered users from the measure to address these concerns. We remain of the view that this would reduce the effectiveness of the measure by encouraging perpetrators to seek out services which allow unregistered users, and therefore recommend the measure applies to all service providers even where there are recognised difficulties such as those with unregistered users. We also emphasise that the measure recommends providers take whatever steps are reasonable for preventing banned users from returning to the service, and that these steps may differ between registered and unregistered users. We also note that as technology advances, the steps a provider may take will also change and that therefore what is deemed to be a reasonable step may also change over time.

Duration of ban – adult users

3.63 In our June 2025 Consultation, we set out three options for the duration of a ban (separately to our position on the treatment of child users):

- Option 1: A permanent ban should be applied in all cases.
- Option 2: Providers should have full discretion to decide the length of the ban.
- Option 3: A permanent ban should be applied as a default, except for in limited exceptional circumstances.

3.64 In the consultation, we proposed taking forward Option 3, that users should be banned permanently as a default, except for in limited exceptional circumstances. In most cases this option would still lead to a permanent ban, thereby preventing users from returning to a service and causing further harm. We considered this would act as a strong deterrent against future perpetrating and more closely reflect the severity of harm than Option 2, yet

[160] Check My Ads Institute response to the June 2025 Consultation, p.10; Ofcom/Marie Collins Foundation Lived Experience Workshop, 3 September 2025, published as response to the June 2025 Additional Safety Measures Consultation, p.5; We Are Survivors response to the June 2025 Consultation, p.7.
[161] British and Irish Law Education Technology Association (BILETA) response to the June 2025 Additional Safety Measures Consultation, p.24.
[162] Middle Tech Coalition response to the June 2025 Consultation, p.6.

would allow for more nuance than Option 1 to ensure proportionality where necessary. We proposed that we would not specify explicitly what may constitute an exceptional circumstance because making these factors known publicly could be exploited by perpetrators to receive a shorter ban.

3.65 Several civil society and government stakeholders including Barnardo's and [✂] agreed with this approach.[163] The Scottish Government said Option 3 reflects the seriousness and impact of CSEA while being proportionate.[164]

3.66 Industry stakeholders highlighted several exceptional circumstances in which they do not see a permanent ban as appropriate. Google LLC raised the example of "misguided activism" as an example of circumstances in which they do not see a permanent ban as appropriate.[165] Snap, Inc., raised the example of meme-style CSAM being shared without malicious intent, or where CSAM was shared to raise awareness or report harm. In these exceptional cases, Snap Inc may issue a warning rather than an immediate permanent ban.[166] LinkedIn said that Ofcom does not account for cases of account takeover, where a bad actor accesses and uses the account of an innocent user without their consent.[167]

3.67 The Marie Collins Foundation argued that any exceptional circumstances should be set by Ofcom rather than being up to service providers and should be very limited.[168] On the other hand, Google LLC supported some level of discretion for providers when imposing a ban but said that service providers will be supportive of lengthy, or indeed permanent, bans.[169]

3.68 Having considered stakeholder feedback, we have decided to proceed with Option 3 in respect of adult users. For adults, a ban should be permanent as a default, but providers can impose a time-limited ban in exceptional circumstances where they consider a permanent ban would not be appropriate. Providers should treat a user as an adult if they do not have reasonable grounds to believe that user is a child. We have decided to take a different approach in relation to children and discuss this in the following section, 'Our approach to children'.

3.69 As stated in our June 2025 Consultation, we expect the exceptional circumstances in which it is not appropriate to apply a permanent ban to an adult user to be extremely limited and rare. We consider this will increase the effectiveness of the measure due to the evidence of the persistent nature of offending and the high rates of reoffending. While official recidivism rates are low, a 2025 study showed that self-reported recidivism rates were at 38.3%, suggesting a high level of unreported and undetected reoffending and a continued risk of harm following the initial offence taking place.[170]

---

[163] Barnardo's response to the June 2025 Consultation, p.14; Check My Ads Institute response to the June 2025 Consultation, p.10; Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p.7; Marie Collins Foundation response to the June 2025 Consultation, p.9; [✂]; Scottish Government response to the June 2025 Consultation, p.3; We Are Survivors response to the June 2025 Consultation, p.7.

[164] Scottish Government response to the June 2025 Consultation, p.3.

[165] Google LLC. response to the June 2025 Consultation, pp.30-31.

[166] Snap, Inc. response to the June 2025 Consultation, p.13.

[167] LinkedIn response to the June 2025 Additional Safety Measures Consultation, pp.8-9.

[168] Marie Collins Foundation response to the June 2025 Consultation, p.9.

[169] Google LLC. response to the June 2025 Consultation, p.30.

[170] Baskurt, S., Babchishin, K., Hilkes, G., and Seto, M., 2025. A meta-analysis of recidivism rates among individuals who commit child sexual exploitation material (CSEM) offending, Aggression and Violent Behaviour, Vol 85.

3.70 We have decided to confirm our position not to explicitly set out the circumstances which may not lead to a permanent ban, due to the risk of aiding perpetrators to circumvent a ban. We acknowledge that this approach creates some regulatory uncertainty, but we consider this to be the best approach for the reasons we have set out in paragraph 3.27. This approach also mitigates the risk that our measure would not capture all the circumstances in which a time-limited ban would be appropriate.

3.71 To ensure bans of a lesser duration effectively protect users from harm, we are recommending that providers must reasonably consider how long it would be appropriate to ban the user for, ensuring that the length reflects the severity of the harm and effectively protects children from harm by deterring the banned user (and other users) from carrying out further harm. This should provide a safeguard against users being banned for an inappropriately short duration.

Our approach to children

3.72 Under UK law, a child who takes a sexualised image or video of themselves has created CSAM and sharing that image or video with another user amounts to an offence. We recognise that despite being an offence, there is a need to take a proportionate and sensitive response in these circumstances to ensure that the child is not unduly penalised or victimised.

3.73 In our June 2025 Consultation, we proposed three options for managing this:

- Option A: Children are not banned, meaning that all children are exempt from the proposed measure.
- Option B: Ban children who commit CSEA harm except in the following exceptional circumstances (which we now refer to as 'extenuating circumstances') where a ban would not be appropriate or proportionate:
  - Child victims and survivors who have been coerced or manipulated into sharing self-generated intimate imagery (SGII);
  - Older children who have sent SGII to another child (or received from another child) as part of a consensual age-appropriate relationship.
- Option C: Ban all children who are identified as having engaged in CSEA and rely on the appeals process for cases where banning may have a disproportionate impact.

3.74 In our June 2025 Consultation, we proposed to reject Option A. It was our view that excluding all children from the measure would reduce its effectiveness due to the risk of harm that can be caused by children who commit CSEA offences and the incentive this would give adult users to present as children (for example by declining to undergo highly effective age assurance) to avoid a sanction.

3.75 We sought views from stakeholders on the options proposed and asked for any alternative views on how we could best address this issue to mitigate the risk of disproportionately banning children while maintaining the measure's effectiveness. We also conducted a workshop with children providing an accessible way for them to comment on our proposals.

Feedback on our proposed options

3.76 Generally, stakeholders welcomed our acknowledgement of the complex circumstances in which children may fall in scope of this measure and agreed with the need for a nuanced

approach which manages the issue sensitively and proportionately.[171] Civil society stakeholders, including the Lucy Faithfull Foundation, the Commissioner Designate for Victims of Crime Northern Ireland and the Online Safety Network outlined the varied contexts in which children may share CSAM and how children can act without malicious intent or without a full understanding or appreciation of the implications or severity of their behaviours.[172] Evidence was provided demonstrating the prevalence of image sharing among children and the increased risks of harm for certain vulnerable groups.[173]

3.77 C3P objected to our use of the word "exceptional" given that this implies that such incidents are rare. It further highlighted the need to encompass the varied scenarios in which children can be coerced or deceived into sharing CSAM.[174]

3.78 The Marie Collins Foundation challenged our use of the term "SGII" stating that it places undue emphasis on the child's behaviour instead of the conduct of adult perpetrators.[175]

3.79 In considering our proposed options, BILETA agreed that Option A would not be appropriate.[176]

3.80 Several stakeholders expressed a preference for Option B, stating this was the most balanced option and best recognised the complexities involved in child image-sharing.[177] Internet Matters said that this reflected the seriousness of CSEA offences while allowing for alternative actions where a permanent ban is not appropriate.[178]

3.81 Stakeholders called for providers to be permitted discretion to make banning decisions on a case-by-case basis. [179] Snap, Inc. said that services should be permitted discretion to determine the exceptional circumstances in which a ban would not be applied as this would enable them to make more informed and context-driven decisions based on user intent and behaviour.[180]

3.82 The National Center for Missing & Exploited Children (NCMEC) supported Option C based on the protection that this approach offered to children.[181] Other stakeholders challenged

---

[171] 5Rights Foundation response to the June 2025 Additional Safety Measures Consultation, pp.11-12; A member of the Philippine Survivor Network response to the June 2025 Consultation, p.20; BILETA response to the June 2025 Consultation, p.25; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.12; Marie Collins Foundation response to the June 2025 Consultation, p.10; NCMEC response to the June 2025 Consultation, p.6; NSPCC response to the June 2025 Consultation, pp.40-45; Ofcom/Marie Collins Foundation Lived Experience Workshop, 3 September 2025, published as response to the June 2025 Consultation, p.5; [✂]; [✂].

[172] Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p.7; Lucy Faithfull Foundation response to the June 2025 Additional Safety Measures Consultation, p.9; Online Safety Network response to the June 2025 Additional Safety Measures Consultation, p.23.

[173] BILETA response to the June 2025 Consultation, p.23; Internet Matters response to the June 2025 Consultation, pp.12-14; IWF response to the June 2025 Consultation, pp.46-47; NSPCC response to the June 2025 Consultation, pp.40-41; Scottish Government response to the June 2025 Consultation, p.3.

[174] C3P response to the June 2025 Consultation, pp.16-17.

[175] Marie Collins Foundation response to the June 2025 Consultation, p.11.

[176] BILETA response to the June 2025 Consultation, pp.25-26.

[177] Barnardo's response to the June 2025 Consultation, pp.14-15; BILETA response to the June 2025 Consultation, p.27; Evangelical Alliance response to the June 2025 Additional Safety Measures Consultation, p.8; Google LLC. response to the June 2025 Consultation, p.30; [✂]; Scottish Government response to the June 2025 Consultation, p.3; [✂].

[178] Internet Matters response to the June 2025 Consultation, p.14.

[179] Snap, Inc. response to the June 2025 Consultation, p.14; [✂].

[180] Snap, Inc. response to the June 2025 Consultation, p.14.

[181] NCMEC response to the June 2025 Consultation, pp.6-7.

this approach, stating that it failed to allow for any discretion to account for the different circumstances in which child users may engage in CSEA and may disproportionately penalise children.[182]

3.83 While there was a general acknowledgement of the need to apply some form of sanction to alert children to the consequences of their behaviours and deter them from causing future harm, many stakeholders raised concerns about a permanent ban for children, including those from industry, civil society, and academia, alongside individual respondents, stating that this was too severe and presented several significant risks that may further exacerbate harm:[183]

- Civil society stakeholders, including the NSPCC, highlighted the emotional impact that a ban could have on victims, including worsening feelings of self-blame.[184]
- The NSPCC and other individual respondents raised concerns that a ban would isolate children from important sources of support, such as family, friends or other online communities.[185] The Lucy Faithfull Foundation stated that social isolation can be a risk factor for children engaging in harmful sexual behaviours and therefore banning may increase the risk of harm to those already vulnerable to perpetrating these kinds of behaviours.[186]
- Several stakeholders from civil society, industry and academia raised concerns that fear of a ban may discourage children from reporting online abuse and prevent children from accessing help and support.[187]
- Finally, concerns were raised by several stakeholders from industry and civil society that a ban may inadvertently push children to less regulated and potentially more dangerous environments which would in turn increase the risk of them being exposed to more extreme and harmful content.[188]

3.84 The children we spoke to as part of the Children's Online Safety Advisory Group (OSAG) workshop supported these views. They stated that cases need to be "handled with protection and proportionality", focused on "safeguarding rather than punishment" and implementing protections that "recognise that young people make mistakes while still

---

[182] Barnardo's response to the June 2025 Consultation, pp.14-15; BILETA response to the June 2025 Consultation, p.27; Internet Matters response to the June 2025 Consultation, p.13; Office of the Children's Commissioner for England response to the June 2025 Additional Safety Measures Consultation, pp. 19-20; [✂].

[183] IWF response to the June 2025 Consultation, pp.46-48; Lucy Faithfull Foundation response to the June 2025 Consultation, p.10; National Police Chief's Council (NPCC) response to the June 2025 Consultation, p.12; Online Safety Network response to the June 2025 Consultation, p.23; [✂]; [✂].

[184] NSPCC response to the June 2025 Consultation, p.42; Ofcom/Marie Collins Foundation Lived Experience Workshop, 3 September 2025, published as response to the June 2025 Consultation, p.5.

[185] Name Withheld 4 response to June 2025 Additional Safety Measures Consultation, p.7; NSPCC response to the June 2025 Consultation, p.42; [✂].

[186] Lucy Faithfull Foundation response to the June 2025 Consultation, p.10.

[187] Barnardo's response to the June 2025 Consultation, p.14; BILETA response to the June 2025 Consultation, p.25; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.13; Google LLC. response to the June 2025 Consultation, p.29; NSPCC response to the June 2025 Consultation, p.43; Online Safety Network response to the June 2025 Consultation, p.23; [✂].

[188] Computer and Communications Industry Association (CCIA) response to the June 2025 Additional Safety Measures Consultation, p.13; IWF response to the June 2025 Consultation, p.47; Lucy Faithfull Foundation response to the June 2025 Consultation, p.10; [✂].

ensuring that uncomfortable material is not there". They also highlighted the importance of signposting to support services, upload prevention, and improved education.[189]

Our decision on our approach to children

3.85 Having considered stakeholders' responses on this issue, we have decided to proceed with Option B as presented in our June 2025 Consultation, meaning that some form of ban should be applied to a user that the provider has reasonable grounds to believe is a child in all cases, except for those where the provider has reasonable grounds to believe that one of the two extenuating circumstances applies. In cases where an extenuating circumstance applies, a service provider may choose not to apply a ban.

3.86 This approach ensures that children who have been a victim of grooming are not disproportionately sanctioned. In response to stakeholder feedback, we want to clarify that the extenuating circumstances include all situations where children have been coerced, pressured, deceived, or bullied – including those who are bullied by another child – into sending CSAM, all of which can amount to a grooming offence. It also ensures that children who have shared consensual images in an age-appropriate relationship are not banned, as we consider this would be disproportionate given the evidence which indicates that, for some children, the sharing of imagery has become an increasingly normal part of their relationships.[190]

3.87 We recognise the unintended implications raised by stakeholders of the use of the word "exceptional" when discussing these circumstances and how this may not accurately reflect the high prevalence of the circumstances we have described. We have therefore opted to use the word "extenuating" to describe these scenarios.[191]

3.88 We also accept the challenges associated with the term "self-generated indecent images (SGII)" and that this is a widely contested term. The term SGII fails to capture the complex nature of child sexual abuse and may unintentionally imply that a child is responsible for their own abuse where these images are created in the context of coercion. In our Code measure, we have used the phrase "CSAM or CSAM proxy of themself". We think this term accurately articulates that the content is abusive material and amounts to CSAM as defined by UK law.

3.89 However, we use the term SGII within our Statements as this is the term currently most used and understood within the online child protection system.

3.90 We have decided to discount Option C as this would impose an overly punitive and disproportionate blanket ban on all children, including victims. We also confirm our decision not to proceed with Option A.

3.91 Having considered stakeholder feedback, we have decided to amend our measure to allow for a more nuanced approach to child users in cases where an extenuating circumstance does not apply. We recognise that there are situations, other than the two extenuating circumstances set out in paragraph 3.12 where children may share or receive CSAM without malicious intent, including where a child may not understand the illegality of their actions

---

[189] Children's Online Safety Advisory Group (OSAG) meeting, 15 January 2026.

[190] An IWF study into the sharing of nude images among children and young people found that for older participants (aged 13-14 years old and over) felt that the requesting and sharing of nude images was a regular and normalised aspect of daily life. IWF and ARU, 2023. Its normal these days. Self-generated child sexual abuse fieldwork findings report. See paragraph 16.65 of the June 2025 Consultation.

[191] We have amended measures ICU D10.3 and ICU H3.6 to reflect this wording.

or may act without recognising the harm that may be caused. Research conducted by Internet Matters found that young people "may lack the knowledge and maturity to appreciate the implications" of their behaviours.[192] This finding was supported in our engagement with the Marie Collins Foundation group and in the evidence submitted by the Digital Resilience in Education team within the Welsh Government.[193]

3.92 We accept that applying a permanent ban in these situations could be disproportionate and may have a significant impact on children. Stakeholder feedback, as set out in paragraph 3.83, highlighted the negative consequences that a permanent ban may have on children including discouraging reporting, creating feelings of guilt and shame and isolating them from their peers and wider social networks.

3.93 Nonetheless, we recognise that in some cases, child users can cause harm and that in these cases, a ban may be proportionate. Recent data by the National Centre for Violence against Women and Girls and Public Protection (NCVPP) shows that in CSEA crime reports in which age was recorded, children, aged 10-17, account for up to 50% of perpetrators.[194] Evidence also attests to the growing risk presented by children involved in COM group activity. COM Groups are online networks of individuals, often under the age of 18, who carry out serious, high-harm offences including child sexual abuse to gain status and notoriety within the network. In the October 2025 Early Threat Identification Notice, it was reported that based on the cases identified, a large proportion of individuals involved in COM group offending are under 18.[195] This demonstrates the severe and intentional harm that can be caused by child users.

3.94 As such, we have decided to recommend that providers should determine the duration of a ban to be applied, where they have reasonable grounds to believe the user is a child and where one of the extenuating circumstances does not apply, or where one of the extenuating circumstances applies but the provider considers it appropriate to impose a ban of some duration. This may be up to and including a permanent ban where they deem this appropriate in light of the harm caused.

3.95 Providers should act reasonably in determining the appropriate duration of a ban and ensure that it reflects the severity of harm to UK users, takes into account whether the user has previously shared CSEA content and whether they were banned, takes into account the potential impact of the duration of the ban on the user, and protects UK users from harm by deterring the banned user and other users from posting CSEA content. In seeking to achieve these outcomes, providers may wish to consider factors such as the age of the user, and any evidence available as to the child user's knowledge or understanding of their actions. We consider permitting providers this discretion will allow for a more nuanced and targeted response, while still reflecting the severity of harm caused by CSEA.

---

[192] Internet Matters, 2023. "It's really easy to go down that path: Young People's experiences of online misogyny and image-based abuse", Online misogyny and image-based abuse research report.
[193] Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.12; Ofcom/Marie Collins Foundation Lived Experience Workshop, 3 September 2025, published as response to the June 2025 Consultation, p.5.
[194] Where perpetrator age was recorded, 50% (n = 40,824 of 81,659) of perpetrators were children aged 10-17 years old, accounting for 57% (n = 17,315) of all non-contact perpetrators and 46% (n = 23,509) of all contact perpetrators. National Centre for Violence against Women and Girls and Public Protection, 2024, 2024 National Analysis of Police Recorded CSAE Crimes Report.
[195] NCA, NPCC and Counter Terrorism Policing, 2025. COM Groups: Early Threat Identification Notice for Safeguarding Professionals.

3.96 In summary, this means that for any user that the provider has reasonable grounds to believe is a child:

a) The provider may decide not to ban that child where one of the following extenuating circumstances exists:

- the child has been groomed into sending SGII, including by coercion, deception, bullying, or manipulation; or[196]
- the child has consensually shared SGII and/or received SGII consensually as part of an age-appropriate relationship.

b) In all other circumstances, the provider should impose a ban of a duration that they reasonably consider to be appropriate, up to and including a permanent ban.

3.97 To determine the duration of a ban for a child user, providers should ensure the duration of the ban:

a) reflects the nature and severity of the harm caused;
b) considers whether the child has previously generated, uploaded, shared or received and viewed CSEA or CSAM content, including whether they were banned from the service for doing so;
c) takes into account the potential impact of the duration of the ban on the child being banned; and
d) protects users, particularly children, from harm by deterring that user, and other users, from sharing CSEA content in the future.

3.98 We expect providers to reflect the need for nuance in how children are treated in their implementation of this measure, but also to account for the need to protect other users on the service. For example, bans should be longer in duration where the harm caused is more severe, and providers may wish to consider bans escalating in duration where a child has repeatedly shared CSAM. However, providers should balance this against the appropriateness of imposing a shorter ban where they consider the child would be particularly negatively impacted by a longer ban. The impact of a ban may differ depending on the child's age, and their culpability or intent in the conduct which has led to the ban.

3.99 An appeals mechanism is available in cases where a user considers a ban to have been incorrectly or disproportionately applied. For children, an upheld appeal may result in a ban being overturned if a provider is satisfied that one of the extenuating circumstances apply, or the duration of a ban being modified where the provider is satisfied that the original length of the ban was inappropriate.

How services may implement this approach to children

3.100 To implement a differential approach to children, service providers should consider whether they have reasonable grounds to believe that a user is a child, and whether one of the specified extenuating circumstances applies. Where an extenuating circumstance does not apply, providers should consider the available evidence to determine an appropriate duration for the ban. It is important to emphasise that this does not mean a provider should determine the precise age of the user, but rather whether they have reasonable grounds to believe the user is a child as opposed to an adult.

[196] When making a judgement as to whether grooming has occurred, see chapter A.6 of the Illegal Content Judgements Guidance.

3.101 Stakeholders sought guidance on how services should assess and determine where extenuating circumstances exist, highlighting the difficulties that providers may have in assessing complex and nuanced circumstances involving child users, in particular given existing barriers to reporting and in the absence of highly effective age assurance.[197] Snap, Inc. sought discretion for providers, but also acknowledged that there can be challenges in making these determinations, in particular, determining where coercion has occurred.[198] The NCA suggested that services may be required to apply a temporary ban on the user while this assessment is conducted to ensure protection for users.[199]

3.102 We accept the challenges that service providers, especially those of small services, may face in implementing a differential approach for child users, in particular the difficulties in considering whether they have reasonable grounds to believe that a user is a child and understanding the context in which content has been shared. The ability to make these considerations will depend on the nature of the case, for example whether there is a user report supporting the action being taken. It is likely that providers of larger services will have more sophisticated means by which to make these determinations, but we consider providers of smaller services will still in many cases have information available to them to reach a view on whether to ban and, if so, for how long. Where a provider does not have reasonable grounds to believe that a user is a child, they should be treated as an adult under this measure.

3.103 We are not recommending providers of user-to-user services implement highly effective age assurance for the purposes of this measure, given the financial impact that this would have on providers of smaller and lower risk services who fall within its scope. Further, the highly effective age assurance that we recommend providers use for the purpose of complying with their safety duties seeks to confirm that users are adults, rather than confirming that they are children. Where a user does not confirm they are an adult, our measures recommend that they should be treated as a child.[200] If we were to recommend providers use highly effective age assurance for the purpose of determining who is a child under this measure, some adult users could choose not to age assure on a service, in which case they would be treated as a child. This could mean they wrongly receive a ban of a lesser duration, or are more likely to convince the provider that one of the extenuating circumstances exists so that they are not banned at all. To mitigate this risk, we instead expect providers to use the information available to them, which may or may not include highly effective age assurance, to consider whether they have reasonable grounds to believe the user is a child, alongside what can be learnt about the circumstances of the case. This information should be used to decide the most appropriate and proportionate response on a case-by-case basis.

---

[197] Barnardo's response to the June 2025 Consultation, p.15; BILETA response to the June 2025 Consultation, p.26; Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p.7; Evangelical Alliance response to the June 2025 Consultation, p.8; [✂]; Google LLC. response to the June 2025 Consultation, p.30; Internet Matters response to the June 2025 Consultation, p.14; [✂; [✂]; NSPCC response to the June 2025 Consultation, p.44; Ofcom/Marie Collins Foundation Lived Experience Workshop, 3 September 2025, published as response to the June 2025 Consultation, p.5; [✂]; We Are Survivors response to the June 2025 Consultation, p.7.

[198] Snap, Inc. response to the June 2025 Consultation, p.14.

[199] NCA response to the June 2025 Consultation, p.18.

[200] Broadly speaking, this means they are prevented from encountering content that is harmful to children, or protected from it (depending on the specific type of content in question).

3.104 We note that as technology develops and adoption of age assurance increases, the information available to providers to make these determinations will change. We therefore expect providers' ability to determine an appropriate sanction will improve over time in line with technological advances.

3.105 Where a provider lacks reasonable grounds to believe the user is a child, or determine the circumstances surrounding the alleged perpetrating, we expect them to ban the user. While we accept that this may at times result in child users being banned, we consider this approach to be proportionate to ensure that those who present a risk of harm do not remain on a service due to an incorrect assumption that a user is a child, or about the context of the harm. The measure allows for users to appeal against a decision to impose a ban, at which point they may provide further information to support a decision to be overturned or the length of a ban to be modified.

Alternative sanctions and wider safeguarding actions

3.106 Some civil society stakeholders proposed the use of alternative sanctions, for example time-limited bans, three-strike policies, reduced functionalities, and educational prompts, instead of a permanent ban. Stakeholders suggested that responses should focus on education and prevention and align to the child's age or age range, and their stage of development.[201]

3.107 We recognise that educational warnings and prompts can provide a teachable moment to children and young people, alerting them to the risks and consequences of their behaviour and deterring them from causing further harm. This view was supported by the children we spoke to who highlighted the importance of education and support,[202] and is reaffirmed in research conducted by Internet Matters and the Centre of Expertise on Child Sexual Abuse (the CSA Centre).[203]

3.108 In response to the feedback we received, we considered whether to recommend providers apply a specific alternative sanction for child users, for example strikes, limited functionalities or educational messages. However, we currently lack evidence as to the effectiveness of each of the proposed alternative sanctions within this context.[204]

3.109 We nevertheless recognise the need for a nuanced approach in relation to children, and explain in paragraphs 3.90-3.94 above why we consider recommending that providers should determine the appropriate length of a ban where they have reasonable grounds to believe the user is a child will support providers to act proportionately.

---

[201] 5Rights Foundation response to the June 2025 Consultation, p.12; Barnardo's response to the June 2025 Consultation, pp.14-15; CCIA response to the June 2025 Consultation, p.13; Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p.7; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.13; Evangelical Alliance response to the June 2025 Consultation, p.8; Google LLC. response to the June 2025 Consultation, p.31; IWF response to the June 2025 Consultation, p.48; Lucy Faithfull Foundation response to the June 2025 Consultation, p.9; Marie Collins Foundation response to the June 2025 Consultation, p.10; Name Withheld 4 response to the June 2025 Consultation, p.7; [✂]; NSPCC response to the June 2025 Consultation, pp.44-45; Online Safety Network response to the June 2025 Consultation, p.23; [✂]; [✂]; [✂]; [✂].

[202] OSAG meeting, 15 January 2026.

[203] Internet Matters, 2023. Preventing 'self-generated' child sexual abuse.

[204] We are aware from stakeholder responses that Lucy Faithfull Foundation hope to do further work in this space in the future. The Illegal Content Codes are iterative and therefore this position may change as and when further evidence is available.

3.110 Service providers may choose to deploy other sanctions, including educational messaging, alongside a ban.[205] We advise that if deploying these messages, service providers are clear of the purpose of the messaging and use them as an opportunity to offer education, advice and reassurance to children and young people. Where appropriate, this may include signposting to relevant support services such as Shore or Report Remove (depending on the context).[206] If choosing to deploy messaging, providers should also consider the published best practice guidance for communicating with children who have been sexually abused, ensuring language is accessible, trauma-informed, and non-victim blaming.

Benefits and effectiveness conclusion

3.111 We conclude that banning is an effective and proportionate option for protecting users from the harm caused by CSEA. We consider this measure will deliver significant benefits and be effective in achieving its aim of disrupting and preventing perpetrators and protecting users from the harm caused by CSEA, while ensuring providers have sufficient flexibility to treat users, in particular children, appropriately on a case-by-case basis.

Impacts and costs

3.112 This section considers the main direct and indirect costs associated with service providers banning users that generate, share or upload CSEA on the service or knowingly receive CSAM, and in preventing these users from returning to the service.

3.113 We expect that the costs of this measure will vary significantly between services. The proposed measure is scalable and flexible, which means costs are likely to depend on several factors including service size, risk of CSEA, and existing systems and processes.

3.114 Where service providers are also in scope of the user sanctions measure, we expect that there may be some cost synergies when implementing both measures. For example, we expect that service providers are likely to develop their CSEA banning policy jointly with any user sanctions policy, so policy development costs could be lower than our estimates.

Direct costs

3.115 In our June 2025 Consultation, we said that a small service provider that had not identified any CSEA in the past might incur some small costs to develop a policy setting out how to implement this measure, and annual costs of updating this. There could also be an additional low cost of communicating the policy to individuals working in content moderation.

3.116 We said that providers of services with a higher risk of CSEA, or who choose to apply time-limited bans when exceptional circumstances apply, may need to set out a more detailed and complex policy, and therefore might incur higher costs.

3.117 We also said that we expected that costs associated with changes to system infrastructure to prevent user return could vary significantly. A small service provider with a low risk of CSEA may be able to rely on their existing user management system and incur little additional cost. Providers of larger and riskier services which need to ban and prevent many users from returning may choose to implement more sophisticated methods such as

[205] See section: User sanctions policy measures.
[206] Report Remove is an online tool managed by Childline allowing children to confidentially report sexual images and videos of themselves and remove them from the internet. Shore is a website run by Lucy Faithfull Foundation that provides a safe space for teenagers worried about their own sexual behaviour to seek support.

automated bans that incur higher costs, but we considered these would be proportionate due to the higher risk of harm.

3.118 We acknowledged there would also be ongoing costs with applying the banning policy which would again vary widely between service providers. For small service providers that identify a limited amount of CSEA, we said that costs were likely to be relatively small. A provider of a larger, riskier service would need to enforce the policy more frequently and therefore incur higher costs but could reduce cost per decision by using an automated service. We generally expected the costs of applying the measure to increase with the risk of CSEA on a service and therefore to be proportionate to the scale of harm.

3.119 In their responses to our June 2025 Consultation, one academic stakeholder, two civil society stakeholders, one government department, and one individual supported our approach to the assessment of direct costs or our impact assessment more generally.[207]

3.120 However, one industry stakeholder and two individual respondents argued that our assessment failed to account for the diversity of the sector and the disproportionate burden this measure would place on providers of smaller services or startups who may not currently operate dedicated moderation systems.[208] Two individual respondents specifically highlighted costs that may be associated in developing or purchasing moderation tools and/or recruiting, training and supporting human moderators.[209]

3.121 Some responses concerned more general expectations about the operation of content moderation systems. This measure is specifically about user banning in response to identified CSEA content and so we do not consider wider points here. There are separate Illegal Content Codes measures that cover broader recommendations on moderation systems.[210]

3.122 We consider that the diversity of the sector is recognised in this measure. Reasonable steps service providers should take to prevent users from returning to the service can vary depending on factors such as the nature and risk profile of the service in question, what is technically feasible, and the resources of the provider. As discussed in paragraph 3.116, we expect that a service provider that identifies a relatively low volume of CSEA content per year could use their existing user management system to manually remove and prevent users from returning, therefore keeping costs relatively low.

3.123 The cost of applying the banning policy is already included in our cost estimates and we acknowledged in our June 2025 Consultation that there will be an additional cost of communicating the policy to individuals working in content moderation, although we expect this will be low for providers of smaller low risk services with a simple policy.

3.124 We have updated our quantified estimates from consultation in line with the latest Office for National Statistics (ONS) data.[211] However, after rounding our estimates, the costs of this measure are largely unchanged from consultation. We estimate that a small service provider that had not identified any CSEA in the past might incur costs of £200 to £300 to

---

[207] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.22; BILETA response to the June 2025 Consultation, p.29; Check My Ads Institute response to the June 2025 Consultation, p.3; Department of Justice for Northern Ireland response to the June 2025 Consultation, p.16; [✂].
[208] [✂]; [✂]; [✂].
[209] Name Withheld 9 response to the June 2025 Additional Safety Measures Consultation, p.3; [✂].
[210] Illegal Content Codes, Measures ICU C1-C8.
[211] See annex 5: Further detail on economic assumptions and analysis.

develop a policy setting out how to implement this measure, and annual costs of around £50 to £80 of updating this.[212] In terms of applying the banning policy, for a small service provider that identifies a limited amount of CSEA we estimate that costs could be around £40 to £90 per account removed.[213]

3.125 Where a provider has reasonable grounds to believe that the user in question also has access to other regulated user-to-user services provided by it, there will also be costs associated with preventing the user from accessing these services.

3.126 Likewise, where a provider has reasonable grounds to believe that the user in question is a child, there will be an additional cost associated with determining the appropriate length of a ban. For a small service provider with a limited amount of CSEA, this cost is likely to be very small, although it may be more significant for larger, higher-risk services.

3.127 None of the responses discussed in the section 'Our reasoning' provided compelling evidence for changing our assessment of costs. Our assessment of the direct impacts remains unchanged.

Indirect costs

3.128 In our June 2025 Consultation, we said this measure could indirectly affect the ongoing costs of implementing some of our existing measures, potentially increasing costs if it resulted in more appeals and training for content moderators. However, where the measure discourages future CSEA content, content moderation costs could be reduced.

3.129 In addition, the banning of users could lead to a reduction in user numbers and therefore revenues for service providers with a high incidence of CSEA. However, we do not place any weight on any loss arising from banning perpetrators. Conversely, the measure could make users feel safer engaging on services and therefore potentially increase user numbers and revenues.[214]

3.130 Indirect costs may also be incurred by service providers who currently use a content proxy rather than making an illegal content judgement on whether content is CSEA. This may lead to services banning users who have breached terms of service but have not uploaded illegal content. To decide how to follow this measure, some providers will need to choose whether to continue using a broad proxy which may lead to the banning of more users than they would like, or whether to modify their systems and processes to make illegal content judgements on whether individual content is CSEA. The approach a provider takes is likely to depend on various factors, such as the number of content proxies they identify and the commercial and reputational incentives on each service.[215]

---

[212] For a description of the labour time assumptions that sit behind these estimates, see paragraphs 16.92 to 16.94 in our June 2025 Consultation.
[213] Based on the assumption that this takes 1.5 hours of a content moderator's time, noting that the actual time required could be higher or lower than this depending on the complexity of the case.
[214] The Welsh Government agreed that there could be positive indirect effects such as increased revenue. Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.14.
[215] We also note any decision on how to follow this measure may be affected by other obligations or expectations on service providers, for example the reporting duty under section 66 of the Act.

Rights impact

Freedom of expression and association

3.131 In our June 2025 Consultation, we acknowledged that this measure would interfere with users' freedoms of expression and association,[216] as users' ability to impart and receive information and connect with others on a service would be restricted where they are banned.[217] We explained our view that any interference would be justified and proportionate, noting the measure's contribution to removing and preventing CSEA content online. We highlighted that the measure would pursue the legitimate aims of the prevention of crime, the protection of health and morals, and the protection of the rights of others.

3.132 We also proposed amending the reporting and complaints measures to recommend that, in response to an appeal, providers may overturn a ban imposed under this measure where the content was correctly identified as CSEA, but one of the identified extenuating circumstances for children applies. We stated that providers may impose a time-limited ban where they determine on appeal that an exceptional circumstance means a permanent ban would be inappropriate.[218] We considered these measures to act as safeguards for users' rights.

Feedback

3.133 In response to our June 2025 Consultation, the Online Safety Network agreed with our assessment of the human rights impact of the measure and highlighted that different considerations should be given to the protections offered to adults and children.[219]

3.134 Other stakeholders raised concerns about the impact on users' right to freedom of expression where users are banned without a criminal conviction having been secured and queried the available redress for incorrect sanctions.[220] Stakeholders also highlighted the need for accuracy and fairness in applying a ban, and the potential censoring of legal speech, with excessive enforcement risking a chilling effect on freedom of expression.[221] [✂].[222]

3.135 Some stakeholders also provided feedback regarding users being banned after receiving CSAM without requesting it or having unintentionally encountered CSAM. Several civil society and industry stakeholders including BILETA, Google LLC, [✂], [✂] and [✂] said that imposing a ban on users who received CSAM without requesting it may not be

---

[216] The right to freedom of association is set out in Article 11 of the European Convention on Human Rights (ECHR). Article 11(2) says that any interference with this right must be prescribed by law, necessary in a democratic society, and that this right may only be restricted by in the interests of national security or public safety, for the prevention of disorder or crime, for the protection of health or morals or for the protection of the rights and freedoms of others.

[217] For the full rights assessment on which we consulted, see paragraphs 16.104 to 16.117 of the Additional Safety Measures consultation document.

[218] We also proposed that providers may reverse a ban where they incorrectly determined content to be illegal, which applies to CSEA content. We have decided to proceed with that proposal (see measure ICU D10.2).

[219] Online Safety Network response to the June 2025 Consultation, pp.24-25.

[220] Cook, J. response to the June 2025 Additional Safety Measures Consultation, p.2; [✂].

[221] Open Rights Group response to the June 2025 Additional Safety Measures Consultation, p.34; [✂].

[222] [✂]; [✂]; [✂].

proportionate, and requested clarity on this point.[223] [224] Friends2Follow requested Ofcom differentiate in the measure where CSAM has not been sought or solicited by the recipient, or refer to receivers as those who "solicit CSAM".[225] A respondent suggested that the measure should focus on users who solicit, request or accept CSAM, so that users who receive it unsolicited are not penalised.[226] The CCIA, [✂], and the Welsh Government also said the draft measure failed to address situations where users accidentally encounter CSEA content and noted that this could deter users from reporting such content.[227]

3.136 Two industry stakeholders also called for more discretion and flexibility in the approach to banning and raised concerns that legal requirements for single-violation sanctions could lead to over-enforcement, resulting in disproportionate impacts on users' rights.[228]

Interference

3.137 As set out in our July 2025 Consultation, we accept that the measure is likely to interfere with users' rights to freedom of expression and freedom of association. There is also a potential impact on service providers' equivalent rights.

3.138 We accept that the interference is particularly severe in relation to adults, in that we are recommending that the default be a permanent ban.

Prescribed by law

3.139 The measure is clear in its recommendation that users who share, generate, or upload CSEA on the service, or receive and view CSAM[229] should be permanently banned from the service on which the content was posted and any other services operated by the provider, subject to certain exceptions. Though we have decided to give providers some discretion in determining whether those exceptions apply, we consider that this does not prevent users from being able to regulate their conduct appropriately. Both the Act and our Illegal Content Codes are published and providers should also refer to their CSEA banning policy in their terms of service, so that users understand when they will be banned. We consider we have given providers sufficient guidance on the objectives of the measure and the factors to consider when applying a ban for it to be sufficiently clear for them to follow the measure.

Necessary for a legitimate aim

3.140 We have explained in the section 'Effectiveness and benefits' the risk of repeat offending and escalation of harm by users who commit CSEA harms. It follows for the prevention of crime, the protection of health and morals, and the protection of the rights of others, it is necessary to take action to prevent repeated conduct by those who have shared, generated, or uploaded CSEA content and CSEA content proxy, and users who have

---

[223] BILETA response to the June 2025 Consultation, pp.24-25; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.12; [✂]; Google LLC. response to the June 2025 Consultation, pp.28-29; IWF response to the June 2025 Consultation, p.44; [✂]; [✂; [✂]; [✂].

[224] Other stakeholders suggested that banning users who receive CSAM may increase the incidence of users being sent CSAM to get them banned. Google LLC. response to the June 2025 Consultation, p.29; Name withheld 4 response to the June 2025 Consultation, p.7. [✂].

[225] Friends2Follow response to the June 2025 Consultation, p.7

[226] [✂].

[227] CCIA response to the June 2025 Consultation, p.12; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.12; [✂].

[228] [✂]; [✂].

[229] Unless they have reported and deleted the content, in which case the provider may determine not to apply a ban.

received and viewed CSAM and CSAM proxy (and have not promptly reported and deleted that content).

3.141 We considered whether this measure is the least restrictive option to achieve the stated aims and conclude that it is, given that other options such as strikes or only removing an account (i.e. not removing all accounts and not taking steps to prevent return) would not as effectively reduce the prevalence of CSEA.

3.142 We consider that banning adult users after one breach is necessary to disrupt offending, addressing the repeated and persistent nature of offending and the escalation of offending behaviours. In the rare exceptional circumstances in which that is not necessary, the measure allows providers to make an exception, and apply a ban of a lesser duration. For these same reasons, we consider it necessary for users to be banned on any regulated service operated by the provider to which it has reasonable grounds to believe the user has access.

3.143 For child users, we consider that a ban from the service remains appropriate in most cases, in order to disrupt offending from an early age, and to appropriately reflect the severity of harm caused by CSEA offences. We recognise that there are circumstances in which a ban would not be appropriate – namely where a child has been groomed into sharing SGII, or where SGII is shared as part of an age-appropriate relationship, and have built these into our measure. In all other cases involving children, we accept that a permanent ban after one breach may significantly impact social development for child users, and that a permanent ban from a service may not appropriately reflect their understanding or culpability for their offending.

3.144 We considered whether it would be appropriate to recommend that users 'may' determine the appropriate duration of a ban where they have reasonable grounds to believe the user is a child, rather than recommending that they 'should' do so. However, we acknowledge that a recommendation that providers 'may' determine the appropriate duration of a ban could still result in providers choosing to permanently ban all children who fall within scope of the measure. For the reasons set out above, we do not consider such a recommendation would sufficiently account for the nuance required when considering the impact on children of a permanent ban.

3.145 We have therefore determined that the least restrictive option to achieve the stated aim in respect of child users is to recommend that providers determine the appropriate duration of a ban where they have reasonable grounds to believe the user is a child. We expect providers to act reasonably when determining what an appropriate length of ban would be, seeking to achieve the factors we set out at paragraph 3.71. We consider this strikes the right balance between protecting users, particularly children, from the severe harm caused by CSEA, while appropriately mitigating the impact on the rights of children under this measure.

3.146 Regarding stakeholder feedback on banning users without a criminal conviction, we acknowledge that it is not possible to identify all perpetrators in the real world and, even in cases where perpetrators can be identified, it may take significant time for criminal investigations and proceedings to be completed. Meanwhile, perpetrators' harmful conduct online may continue unchecked, and a lack of sanctions for their actions online may encourage others to behave similarly. Online harms can proliferate quickly, with perpetrators able to act with relative anonymity. Swift and thorough action is needed to

address such harms, and we consider that the considerable harm caused by CSEA offending serves to heighten the need for such action to be taken.

Proportionate

3.147 We have carefully weighed up the competing rights of those affected by this measure. We acknowledge that the impact on those banned may be significant. In particular, the loss of the ability to express themselves and associate freely online can have significant impacts on a user's reputation, their livelihood, their political life and their ability to maintain relationships with family and friends. A permanent ban, which may have effect across multiple services, is particularly intrusive.

3.148 We also acknowledge that in order to give effect to the measure, the service provider must make judgments based on information which may be incomplete. In particular, in relation to children, evidence of coercion or bullying may not be available to the service provider. It may only exist offline, or have been deleted, or be on a different service. We acknowledge stakeholder feedback on the difficulty of establishing whether a defence applies. There is a risk that users will be banned who are not culpable. This risk exists in relation to both adults and children.

3.149 However, in our judgment, the severity of the harm of CSEA, as set out in paragraphs 3.37-3.42 warrants this. CSEA offences have been designated by UK Parliament as priority offences, reflecting their severity and the substantial public interest in measures which are intended to reduce the prevalence and dissemination of CSEA content online. CSEA offences carry a high rate of recidivism, and evidence shows that perpetrators will seek to evade bans to continue offending.[230] CSEA offences are severe, and can have a significant impact on victims and survivors. Offending behaviour can become entrenched and normalised online. There is broad international consensus on the need to tackle CSEA harm online.

3.150 Earlier this year, the UK Government announced that it will introduce a restriction on social media services offering their services to under-16s using a new power to make secondary legislation under the Children's Wellbeing and Schools Act 2026. It also announced that a wider range of services would need to restrict under 16s from accessing certain features such as livestreaming and communicating with strangers, and restrict them by default for 16- and 17-year-olds.[231] We recognise that the introduction of these restrictions means that fewer children should be able to access social media services within the scope of this measure. However, children aged 16-17 will still be able to access social media services in scope of this measure and protecting older children from the harm caused by CSEA remains an important, legitimate aim of this measure. Younger children will still be able to access some services in scope of this measure, such as direct messaging services on which there is still a risk of harm caused by CSEA. The intended protection of children from this measure also extends beyond whether children can see CSEA content and also concerns protecting children who are subject to CSEA which is shared online. Finally, the benefits conferred by this measure are intended to protect not only children from the risk of harm presented by

[230] Ofcom (2024), Illegal Harms Register of Risks, p.75.
[231] UK Government (2026), Social media to be banned for under-16s in landmark government move to give kids their childhood back - GOV.UK, UK Government, 2026. Growing up in the online world: government response (July 2026).

CSEA, but also adults, who will be unaffected by the Government's proposed social media ban.

3.151 We acknowledge the stakeholder feedback regarding users who have not solicited CSAM, who may have been sent such content without their knowledge, or may have unintentionally encountered the content. Knowingly making or possessing CSAM, both of which may arise where a user receives CSAM, are offences as a matter of law in the UK (unless a defence applies).[232] We therefore consider it important that "receiving" CSAM is included in the measure, to ensure that both perpetrators who receive CSAM from children they have coerced and/or groomed and adult perpetrators who receive CSAM from other perpetrators receive a ban.

3.152 To avoid users being banned in circumstances where they do not have a guilty state of mind, we have decided that providers should have discretion not to ban users where they receive and view CSAM, but promptly report and delete the content in question (as discussed in the section titled 'How this measure works'). A user who reports the content to the police rather than to the service provider will be able to show evidence of this to the provider in appealing the ban. This will enable providers to avoid banning users who have inadvertently or unwillingly encountered CSAM on a service, ensuring their rights are not disproportionately interfered with. However, we are recommending that providers may still choose to ban a user who has promptly reported and deleted CSAM which they have received and viewed, as we recognise it may be appropriate to ban a user in these circumstances. We consider this approach to be proportionate in ensuring the policy achieves its aims while mitigating any impact on the human rights of users who fall within this exception.

3.153 We agree with stakeholder feedback regarding the importance of flexibility to mitigate any interference with users' rights. Although we consider that, for adults, a ban should be permanent by default, we also acknowledge that there may be exceptional circumstances where a permanent ban may not be appropriate. In these circumstances, providers will have the flexibility to implement a ban of a shorter period if they consider it appropriate, ensuring that any impact on their rights is proportionate.

3.154 Reflecting their reduced culpability, we have made specific provision to protect users who are children. As set out in our 'Our approach to children' section, we are recommending that providers should determine the appropriate duration of a ban where they have reasonable grounds to believe the user is a child. This will mean that providers can take into account any evidence of the circumstances that led to the sharing of the content, to ensure that the duration of the ban is appropriate and that any impact on the user's rights is proportionate in those circumstances. Additionally, where there is evidence that the child was groomed, bullied or coerced, or that they were acting within an age-appropriate consensual relationship, we have provided for a specific exception where providers may not ban at all, as we consider it will almost always be disproportionate to permanently ban a child in these cases.

3.155 We recognise the difficulties providers may face in establishing whether they have reasonable grounds to believe a user is a child, and in assessing the specific circumstances of each case. Our measure recommends that, where providers do not have reasonable grounds to believe a user is a child, they should be treated as an adult and a permanent ban

[232] See e.g. *R v Smith* (2002). EWCA Crim 683.

should apply (unless an exceptional circumstance warrants a shorter ban). We accept that this will lead to some children being banned permanently where they may not have been if the provider had access to further information, but consider the interference to which this would give rise to be proportionate in order to achieve the overarching aim of protecting users, in particular children, from the considerable harm caused by CSEA. To mitigate this impact and ensure the measure is proportionate, users will be able to appeal against a ban and provide further information to the provider to seek to have the ban overturned where an extenuating circumstance applies, or to be subject to a shorter ban in other circumstances.

3.156 We also agree with the need for accuracy and fairness in assessing CSEA content leading to a user being banned from a service to ensure that any interference with users' rights is justified and proportionate. We recognise the risk that users are banned for sharing content that the provider incorrectly determines to be CSEA. Where applicable, measures ICU C3 (Setting internal content policies), ICU C4 (Setting performance targets, which include accuracy performance targets), ICU C6 (resourcing), ICU C7 (Provision of training and materials to individuals working in content moderation (non-volunteers) and ICU C8 (Provision of materials to volunteers) all help to secure that a provider makes accurate and fair judgments. Measure ICU C4A will also act as a safeguard, allowing a provider to reverse a decision that content is illegal content or illegal content proxy, other than in relation to a content appeal. For more detail on this measure, please see chapter 9: Deployment of proactive technology to detect image-based CSAM. For providers recommended to deploy perceptual hash-matching, measures ICU C9.6 to ICU C9.18 are further safeguards. We acknowledge that some image-based CSAM will be detected through proactive technology, in line with measure ICU PT1. The method through which the provider has detected CSEA or CSAM content does not impact the application of the user banning measure and, where image-based CSAM is detected through the use of proactive technology, measure ICU C4A will also act as a safeguard. Our Illegal Content Judgements Guidance is also intended to support service providers in understanding how to make judgements about illegal content.

3.157 Other measures also act as safeguards for users' rights, including the reporting and complaints measures. We have amended the reporting and complaints measures to permit providers to shorten the duration of a ban imposed under this measure, where the provider correctly determined that an adult posted CSAM, but an exceptional circumstance applies, or where the provider correctly determined that a child posted or received CSAM but a shorter ban is warranted in the circumstances. They may also overturn a ban where the content was incorrectly identified as CSEA, where the user was banned for receiving and viewing CSAM, and on appeal provides evidence that they promptly reported and deleted the content, or where the user is a child and one of the extenuating circumstances applies.

3.158 As noted above, although it will be difficult in practice for service providers to prevent all banned users from returning to the service, we accept that this measure amounts to a significant interference with the freedom of expression and association of banned users, particularly where they are banned permanently across multiple services. However, we note that the measure does not amount to a ban on use of the internet generally or on users publishing lawfully. Permanently banned users will continue to be able to express themselves and associate in other ways, including in person, in the press and media and on unregulated services. Their access is restricted only by a provider whose service they have attempted to use (or succeeded in using) unlawfully.

3.159 Given the points discussed in the previous paragraphs, we are satisfied that any interference with users' or providers' rights arising from the measure is proportionate to the legitimate aims pursued.

Privacy

3.160 In our June 2025 Consultation, we acknowledged that following this measure would require providers to review and process users' personal information at different stages of the banning process, and so would interfere with users' rights to privacy, which includes the right to establish and develop relationships with others. We considered that any interference would be necessary and proportionate to the measure's legitimate aims due to the severity of the harm caused by CSEA and the effectiveness of the measure in mitigating that harm.[233]

Feedback

3.161 We received comments from the ICO and BILETA regarding the impact on data protection and users' right to privacy. Both stakeholders highlighted the impact of the measure on data protection, with BILETA stating this would be particularly important for recognising and banning connected accounts. BILETA raised that this increased processing of users' data was likely to be more significant given the use of age assurance on services hosting adult content. The ICO also commented on the possibility of providers using solely automated tools to ban users, which would engage Article 22 of the UK GDPR.[234] The ICO stated that there could be unfair processing of personal data if users are not informed of the potential for a ban to occur and noted that the measure does not recommend that providers consider the differing impact of a ban on different users to ensure it is fair. The ICO also highlighted that service providers should aim to minimise the quantity of personal data processed in implementing the measure and referenced their Children's Code for guidance on the processing of children's data.[235] [236]

3.162 [✂].[237]

3.163 The ICO also commented on the privacy safeguards for the measure, suggesting that further elements of the draft codes should be included as privacy safeguards. As set out in the introductory chapter in annex 1: Further stakeholder responses, we consider that the privacy safeguards set out in the Illegal Content Codes for this measure encompass all those relevant to safeguarding users' privacy.

Interference

3.164 We acknowledge that the measure will lead to an interference with users' privacy and engage data protection law where providers have reviewed content uploaded by the user which they determine to be CSEA or CSEA content proxy and then process the user's personal information for the purpose of applying and enforcing a ban (including preventing users from returning to the service). We recognise that providers will also process a user's

---

[233] Ofcom (2025), June 2025 Additional Safety Measures Consultation, paragraphs 16.114 to 16.117.
[234] ICO (n.d.), Rights related to automated decision making including profiling.
[235] BILETA response to the June 2025 Consultation, p.25; ICO response to the June 2025 Additional Safety Measures Consultation, pp.14, 20-24.
[236] In the context of preventing return to a service after being banned, in some instances this may also involve 'storage and access technologies' which engage the requirements set out in Regulation 6 of the Privacy and Electronic Communications Regulations (PECR). The ICO has also provided guidance on this matter. ICO, 2026. Guidance on the use of storage and access technologies.
[237] [✂].

personal information where they are considering whether there are reasonable grounds to believe a user is a child, and whether an extenuating circumstance exists. Providers will also need to process personal information when reviewing appeals against a ban, and when taking steps to ensure that banned users do not circumvent the ban and return to the service.

In accordance with law

3.165 We explain at paragraph 3.120 why our measures satisfy the requirement that any interference be 'prescribed by law'. For the same reasons we consider that the measures are 'in accordance with law' for the purpose of Article 8. We note that data protection law, in particular the principle of data minimisation, constrains providers' ability to process personal data without a lawful basis.

Necessary for a legitimate aim

3.166 As set out in paragraph 3.121, the measure pursues the legitimate aims of the prevention of crime, the protection of health and morals, and the protection of the rights of others. For these reasons, it is necessary to take action to prevent repeated conduct by those who have shared, generated, or uploaded CSEA content and CSEA content proxy, and users who have received and viewed CSAM and CSAM proxy (and have not reported and deleted that content).

Proportionate

3.167 We are satisfied that this measure does not recommend that providers collect or process any more personal information than is necessary to impose a ban on a user. In particular, we are not recommending that providers proactively seek information about whether a user is a child, or whether an extenuating or exceptional circumstance exists. Rather, we are expecting providers to use the information already available to them. That said, providers may choose to collect and process further personal information to inform their decision on banning a user. In all cases, providers should be using this information for the purposes of imposing an appropriate ban. Any interference with a user's privacy arising from the consideration of the circumstances of the posting or receiving of CSEA will therefore need to be balanced against the corresponding benefit of ensuring that the impact on the user's freedom of expression and association is proportionate. In carrying out these actions, providers will have to comply with their data protection obligations, including data minimisation, which will contribute to limiting any interference with users' privacy.

3.168 The measure also includes safeguards for users' right to privacy, including content appeals and complaints. Where appropriate, providers must overturn or may amend the duration of a ban implemented against a user.

3.169 As set out in the introductory chapter in annex 1: Further stakeholder responses, we acknowledge the importance of the measures highlighted by the ICO to help providers comply with data protection law, and note that providers remain subject to applicable data protection law.

3.170 On this basis, we consider any interference with privacy arising from this measure to be proportionate to the legitimate aims connected to protecting users from the severe harm caused by CSEA.

Who this measure applies to

3.171 In our June 2025 Consultation, we proposed that this measure should apply to providers of all user-to-user services to offer the greatest protection to victims and survivors. [✂].[238] Applying this measure to all user-to-user services therefore reduces the risk of the measure inadvertently making it so perpetrators are more able to cause harm on out-of-scope services. It further recognises the potential risk of CSEA on all user-to-user services, including those where this material has not previously been identified. We also expect providers of end-to-end encrypted user-to-user services to ban users where they have determined that the user has committed a CSEA offence, in accordance with this measure.

3.172 Three civil society stakeholders and a government body supported our proposal to apply this measure to all user-to-user services, recognising the importance of this in maximising effectiveness.[239]

3.173 We recognise the costs associated with this measure and the need to ensure that it is proportionate for services of all sizes. The measure gives service providers some discretion over the technical implementation of the ban, so they have flexibility to apply the measure in the most cost-effective way, minimising costs for providers of small, low risk services. Given the severity of CSEA, the impact that CSEA harms can have on victims and the effectiveness of banning, we are of the view that this measure is proportionate even for providers of low risk services where bans will be infrequently required.

3.174 We have therefore decided to recommend this measure applies to all user-to-user services.[240] We are not applying an equivalent measure to search service providers, but we note such providers are in scope of measure ICS F2 (Provision of CSAM warning).

Conclusion

3.175 There is a strong case for taking action to tackle CSEA harm given the persistent nature of CSEA offending, the demonstrable escalation of harm over time, and the severe impact of this offending on victims and survivors.

3.176 In designing this measure, we have recognised that being banned interferes with a user's rights. We have protected against this impact in several ways, including by acknowledging that there will be exceptional circumstances where a permanent ban may not be appropriate, and by amending the reporting and complaints measures to permit providers to shorten the duration of a ban imposed under this measure (see 'Amendments to the Codes').

3.177 We also account for cases where the provider has reasonable grounds to believe that the user is a child. We recognise that this measure may be impacted by the Government's proposed new legislation, but are of the view that the measure continues to present significant benefits in protecting UK users from harm. In circumstances where the provider has reasonable grounds to believe the user is a child, the provider should reasonably determine the appropriate duration of a ban, up to and including a permanent ban. We also

---

[238] [✂]

[239] 5Rights response to the June 2025 Consultation, p.11; C3P response to the June 2025 Consultation, p.16; Online Safety Network response to the June 2025 Consultation, p.23; Welsh Government response to the June 2025 Consultation, p.13.

[240] We have taken a different approach to the User Sanctions measures - see 'Who these measures apply to' in the following section.

recommend that in cases where a child has shared images consensually within an age-appropriate relationship, or where they have been coerced or grooming into sharing SGII, a provider may choose not to apply a ban. In all circumstances, our measure does not stop providers from employing alternative interventions alongside a ban, such as deterrence messaging or educational prompts.

3.178 Given the benefits associated with this measure, we consider the costs impacts on service providers to be proportionate, and that any interference with the rights of users arising from this measure pursues legitimate aims, is prescribed by law and is necessary in a democratic society.

3.179 We conclude that this measure will deliver significant benefits in reducing the risk of harm caused by online CSEA and deliver protections to children. As such, we have decided to proceed with this measure as outlined in paragraphs 3.10-3.15. This measure is part of our Illegal Content Codes of Practice for CSEA and is referred to as ICU H3.

Measures on user sanctions policy

3.180 In the June 2025 Consultation, we proposed two new measures in our Illegal Content Codes and Protection of Children Code of Practice for user-to-user services (Protection of Children Code) that said user-to-user service providers should prepare and apply a sanctions policy in respect of UK users who generate, upload, or share relevant content, with the objective of preventing future dissemination of relevant content.[241]

3.181 Where we use the term 'relevant content' throughout this section, we refer to illegal content, illegal content proxy, the specific kinds of relevant content harmful to children that are prohibited on a service and content harmful to children proxy.

3.182 In our June 2025 Consultation, we said that the policy should set out the sanctions the service provider may apply to users. When setting its policy on when to apply a sanction and the appropriate sanction for relevant harms, we said the provider should have regard to:

a) the severity of potential harm to UK users (UK child users in the case of the sanctions policy for content harmful to children) if they encounter the relevant content on the service;
b) whether the user has previously generated, uploaded, or shared relevant content, including whether they were sanctioned for doing so; and
c) the potential impact of the type and duration of the sanction on the user being sanctioned.

3.183 We proposed that the user sanctions measure ICU H2 in the Illegal Content Codes should apply to providers of all user-to-user services.

3.184 We proposed that the user sanctions measure PCU H2 in the Protection of Children user-to-user Code should apply to providers of user-to-user services that are likely to be accessed by children and that prohibit one or more specific kinds of primary priority content, priority content, and/or non-designated content. In relation to content that is harmful to children, providers should prepare and apply a user sanctions policy for the harms that they prohibit.

[241] Illegal Content Codes, Measure ICU H2; Protection of Children Code of Practice for user-to-user services (Protection of Children Code), Measure PCU H2.

3.185 We said our objective was that providers have measures in place that reduce the future dissemination of relevant content by influencing user behaviour, including by reducing the likelihood of the user posting further relevant content.

3.186 Our Illegal Content Codes also include measures on:

a) banning users who generate, upload, or share CSEA (or CSEA content proxy) to the service, and those who knowingly receive child sexual abuse material (CSAM); and
b) removing accounts that a provider has reasonable grounds to infer is operated by or on behalf of proscribed organisations.[242]

3.187 The illegal content user sanctions measure therefore does not apply in these situations.

3.188 As set out in the introduction to this chapter, we also proposed amendments to several existing measures to support transparent and effective sanction policies.

Our decision

3.189 We have decided to broadly confirm the user sanctions measures we proposed in the June 2025 Consultation. We set out our reasons for doing so in the next section. We have decided, based on our cost and proportionality assessments, that the measures will not apply to providers of services that are small and single or low risk. See the sub-section 'Who these measures apply to' for our full reasoning.

3.190 Our measures now set out that:

a) Providers of services that are large and/or multi-risk for illegal harms should prepare and apply a sanctions policy in respect of UK users who generate, upload, or share illegal content and/or illegal content proxy, with the objective of preventing future dissemination of illegal content.
b) Providers of services that are large and/or multi-risk for content harmful to children, are likely to be accessed by children, and prohibit one or more specific kinds of primary priority content, priority content and/or non-designated content, should prepare and apply a sanctions policy in respect of UK users who generate, upload, or share content harmful to children and/or content that is harmful to children proxy.

3.191 The full text of the measures can be found in our Illegal Content and Protection of Children Codes of Practice and are referred to as ICU H2 and PCU H2 respectively. The measures are part of our Codes of Practice for user-to-user services on terrorism, other duties, and Protection of Children.

3.192 We have also decided to change which services are in scope of the amendments to the reporting and complaints and terms of service measures. We have also made some further amendments to the drafting of the measures to better reflect our policy decisions. Our reasoning is set out in full in the section 'Amendments to the Codes to reflect user sanctions measures'.

How these measures work

3.193 In-scope service providers should prepare and apply a sanctions policy in respect of UK users (regardless of whether they hold an account on the service) who generate, upload, or

[242] Illegal Content Codes, Measures ICU H1 and ICU H3.

share relevant content, with the objective of preventing future dissemination of relevant content.

3.194 As set out in our June 2025 Consultation, we recognise that providers may choose to design their sanctions policies to apply globally. Providers may therefore incorporate these measures within a global policy that also provides for sanctions to be imposed on non-UK users.

3.195 Providers should implement the type of policy that is appropriate for their service, based on insights such as their service type and risk factors. Providers have the flexibility to determine which sanctions to apply. In drafting their policy and determining appropriate sanctions, we recommend providers have regard to three factors, which we explain in more detail in paragraphs 3.201 to 3.220.

3.196 The policy may set out that in certain circumstances, no sanction should be applied (however, action against the content itself must still be taken to comply with measures ICU C2 / PCU C2).

3.197 The user sanctions policy should be applied once a provider has reviewed and assessed the content to be relevant content and has taken a content moderation action. We expect the user sanctions policy to complement providers' existing content moderation processes, and the applicable content moderation measures we have set out in the Illegal Content and Protection of Children Codes.[243]

3.198 Where it is not currently technically feasible for a provider to take content moderation action, providers may apply a user sanction where doing so would prevent the future dissemination of relevant content.[244]

Range of sanctions

3.199 There are a range of actions a service provider could take once it becomes aware that a user has generated, uploaded, or shared relevant content. As stated in paragraph 3.194, the flexibility of these measures allows providers to select the appropriate sanction to prevent future dissemination of relevant content.

3.200 For example, for less severe contraventions, providers may wish to consider providing warnings. In response to more severe harms and/or repeat contraventions, providers may decide to restrict or remove a user's access to the service.

3.201 In Table 3.1, we set out a non-exhaustive list of sanctions that a provider may impose on a user once they are aware that the user has generated, uploaded, or shared relevant content. A provider may take more than one of these actions in response to a single contravention. They may also view different sanctions as appropriate for different contraventions, as set out from paragraph 3.201.

Table 3.1: Examples of sanctions and their description

Sanction	Description
Warnings	

[243] Illegal Content Codes, Measures ICU C1-C2; Protection of Children Code, Measures PCU C1-C2.
[244] Protection of Children Code, Measure PCU C2 outlines other possible actions when content takedown is not technically feasible for categories of content harmful to children.

Sanction	Description
Information	Providers may give users information (for example, about relevant sections of community guidelines and terms of service) to ensure they understand that they should not generate, upload or share the relevant content again.
Educational prompts	Providers may give users educational information to improve users' understanding of relevant topics, for example, where a user has shared suicide, self-harm and/or eating disorder content, to ensure they understand that they should not generate, upload or share the relevant content again.
Strike	Providers may consider the use of a strike system (and how long the strike remains on the user's account), a strike being part of a progressive discipline system that may lead to further penalties or sanctions if the number of accumulated strikes reaches a certain threshold.
Account restrictions	
Preventing access to groups and channels	Providers may prevent a user from accessing groups and channels.
Suspend functionality	Providers may restrict a user's access to functionalities. For example, a provider may temporarily/permanently prevent a user from creating new groups, sending or forwarding messages, creating new chats with users they have not interacted with before, or uploading new content.
Banning	Providers may ban a user, removing the user's access to the service. Providers may additionally block identifiers associated with a specific account from creating new accounts.

Factors providers should have regard to when setting their policy

3.202 Providers should have regard to at least the following three factors when setting their sanctions policy:

a) the severity of potential harm to UK users (and in the case of Measure PCU H2: sanctions policy in respect of users who generate, upload, or share content harmful to children that is prohibited on the service, UK child users) if they encounter the content identified to be relevant content;
b) whether the user has previously generated, uploaded, or shared relevant content, including whether they were sanctioned for doing so; and
c) the potential impact of the type and duration of sanction on the user being sanctioned.

3.203 We consider that having regard to these factors will help providers appropriately balance the need to protect UK users from harms associated with relevant content, and users' rights to freedom of expression, freedom of association, and privacy.

3.204 Providers may choose to take additional factors and context into consideration when setting their sanctions policy.

3.205 The measures' flexibility also allows for providers to have regard to the factors when making individual user sanction decisions. We consider that providing this flexibility will

allow providers to set and apply a sanctions policy in a way that is appropriate for their service.

3.206 We will now examine relevant considerations when applying each of the three factors.

Severity of potential harm to UK users if they encounter the relevant content

3.207 When providers consider the severity of harms, some will be considered more severe than others and therefore warrant a stronger sanction. We highlighted several providers already carrying out this practice in our June 2025 Consultation.[245]

3.208 Providers should consider the findings of their illegal harms risk assessment and (where applicable) their children's risk assessment when considering the severity of potential harm. These findings can help providers to carefully consider the nature and severity of harm, including the number of individuals that could be impacted, the impact on the individuals affected including how user base demographics affect the way in which users experience harm on the service, and any characteristics of the service that may increase the impact of harm.[246]

3.209 When deciding whether a sanction that restricts a user's access to a service is warranted, providers should have regard to the severity of the potential harm for which the user is being sanctioned.

Content identified to be illegal content and/or illegal content proxy

3.210 The UK Government used 'severity' as one of the three factors to determine its list of priority illegal offences when setting out priority offences under the Act. Therefore, when considering severity, providers should consider whether the content in question is priority illegal content or priority illegal content proxy.

Content identified to be content harmful to children and/or content harmful to children proxy

3.211 As set out in the Act, all children should be prevented from encountering primary priority content, and children in age groups judged to be at risk of harm should be protected from encountering priority content and non-designated content.[247] In the Children's Register of Risks (Children's Register), we set out the impacts of children encountering harmful content. These can be severe and, in some cases, fatal.[248] Even within certain harms, there may be different degrees of severity. For example, our Children's Register and 2021 joint report with the Alan Turing Institute highlight how broad priority content categories such as abuse and hate can be, and the difficulty of ascertaining harm in certain situations.[249]

3.212 Therefore, as with illegal content, we consider that the severity of potential harm to child users should be something providers consider when determining appropriate sanctions, as some forms of content harmful to children may warrant harsher or more restrictive sanctions.

---

[245] Meta response to the 2022 Call for Evidence: First phase of online safety regulation (2022 Call for Evidence), pp.28-29; Ukie response to the November 2023 Consultation on Protecting People from Illegal Harms Online (November 2023 Consultation), pp.17-18.

[246] See the Illegal Harms Risk Assessment Guidance and Risk Profiles and the Children's Risk Assessment Guidance and Children's Risk Profiles for further information on assessing severity of harm.

[247] Section 12 (3) of the Act.

[248] Ofcom (2025), Protection of Children's Register of Risks (Protection of Children Register), p.208, 211, 217, 219.

[249] Ofcom and the Alan Turing Institute, 2021. Understanding online hate: VSP Regulation and the broader context; PoC Register, pp.111-145.

Whether the user has previously generated, uploaded, or shared relevant content, including whether they were sanctioned for doing so

3.213 Providers should have regard to whether the user has previously generated, uploaded, or shared relevant content, including whether they were sanctioned for doing so. We consider that it may be appropriate to sanction repeat breaches more severely than a first breach as it is likely a user would have already been made aware that they have previously posted prohibited content. Industry evidence shows that some services use a strike system to sanction users and track repeat offenders.[250]

3.214 Providers may want to consider whether a repeat contravention relates to the same type of content, and whether it would be appropriate to impose escalating sanctions where this is the case. For example, they may want to consider the appropriate action to take if a user is found to have shared eating disorder content for the third time after a warning was sent the previous two times such content was shared.

3.215 Providers may also want to consider for how long any previous sanctions should remain relevant to considering sanctions for further instances of posting relevant content. We understand that some service providers take the age of previous sanctions into account. For example, strikes on TikTok accounts expire after 90 days and will no longer be taken into consideration for a permanent account ban.[251] Similarly, Meta does not count strikes on violating content posted over 90 days earlier for most violations, or over four years for severe violations.[252]

Potential impact of the type and duration of the sanction on the user being sanctioned

3.216 When setting their policy, providers should consider the potential impact of the type and duration of the sanction on the user being sanctioned.[253] We understand that the potential impact on the user may vary depending on the service, as some users may find the features or content of one service more essential than another.

3.217 Different sanctions will affect how users use the service in different ways and therefore have different impacts. For example, long-term sanctions that restrict accounts are likely to have a greater impact on sanctioned users. Providers should therefore consider whether the impact is appropriate given the harm and the user being sanctioned and take this into account when setting the type and duration of a sanction.

3.218 We are aware that users may post harmful content without malicious intent, and that in some cases (such as categories of content harmful to children like eating disorder content) the harmful content could be an indication of vulnerability.[254] Providers may therefore wish to consider the nature of the relevant harm when considering the potential impact of the sanction on the user and take a nuanced approach if they deem it appropriate.

---

[250] Google LLC. response to the 2022 Call for Evidence, p.29; Match Group response to the Call for evidence: Second phase of online safety regulation, p.13; Meta, 2025, Counting Strikes; Snap Inc. response to the November 2023 Consultation, p.25.
[251] TikTok, 2026. Content violations and bans.
[252] Meta, 2025. Counting Strikes.
[253] We note that providers have a duty under Section 22(2) of the Act, when deciding on and implementing safety measures and policies, to have particular regard to the importance of protecting users' right to freedom of expression within the law.
[254] Ofcom (2025), Guidance on Content Harmful to Children, p.32.

3.219 Service providers in scope of measure PCU F4 (Signposting children to support when they post harmful content) should signpost children to resources when they post suicide, self-harm, eating disorder or bullying content.[255]

3.220 Some users, particularly children, may be less aware of the impact of their online behaviour. Providers may therefore also consider whether they have reasonable grounds to infer the user is a child when assessing the potential impact of a sanction. For example, providers may take this into consideration when deciding if banning is the most appropriate sanction. In our measure exploring banning for CSEA offences, we further explore the impact of banning children from accessing services and the importance of considering their experiences when setting sanctions (see section "How services may implement this approach to children ").

3.221 The measures do not recommend that providers implement highly effective age assurance. We instead expect service providers to use the information available to them – which may or may not include highly effective age assurance – to consider the age of the user.

Review of the sanctions policy

3.222 Providers should regularly review and, where appropriate, update their sanctions policy to ensure it remains fit for purpose. In doing so, providers should take into account the findings of their most recent illegal harms risk assessment and (where applicable) their children's risk assessment.

3.223 Providers should also take account of any matters arising from appeals they have determined during the period covered by the review that related to users being given a warning, suspended, banned, or in any other way restricted from using the service. For example, if a provider upholds several appeals relating to sanctions for a particular harm, they may consider whether the relevant part of their policy should be reviewed and updated.

3.224 We will now outline our reasoning for recommending these measures. In doing so, we will review and respond to the stakeholder feedback received in response to our June 2025 Consultation that influenced our decision.

Our reasoning

Effectiveness and benefits

3.225 We consider that there is an ongoing risk of harm to users where providers take content moderation action but do not consider whether to issue a sanction against the user who disseminated the relevant content. This means users may continue generating, uploading, and sharing relevant content, and could mean there is little deterrent for users to stop posting illegal content or content harmful to children.[256]

3.226 Evidence in our Illegal Harms Register demonstrates that, for certain kinds of illegal harm, such as incitement to violence and harassment, there is a risk that providers do not seek to,

[255] Services likely to be accessed by children that (a) are a large service, (b) are at medium or high risk of bullying content, eating disorder content, self-harm content or suicide content, and (c) where a child accessible part of the service has posting content functionality should signpost children to signpost children to support at important points in the user journey.

[256] For existing evidence on repeated sharing, see Guidance on Content Harmful to Children, p.48 and Illegal Harms Register of Risks (Illegal Harms Register), pp.125, 128, 146, 156, 181.

or are ineffective in, preventing banned users from returning to the service after content is taken down.[257]

3.227 Evidence and our analysis suggest that well-calibrated sanctions policies can be effective at reducing the dissemination of relevant content online:

a) There is evidence to suggest warnings, including strikes, are effective at reducing reoffending.[258] [259]
b) Evidence suggests that offenders often repeatedly engage in the same type of conduct online, violating the same policy or same feature. In this context, suspending functionalities for users can disrupt offending online.[260]
c) Further, there is evidence that better communication from services regarding the reasons for content removal and sanctions leads to increased perceptions of fairness and reduces likelihood of reoffending.[261]
d) Responses to the May 2024 Consultation on Protecting Children from Harms Online (May 2024 Consultation) suggested that providers should ensure there are clear consequences for users who consistently break the rules.[262]

3.228 As set out in 'How these measures work', on services where it is not technically feasible for providers to action relevant content, user sanctions can provide an effective means of preventing the future dissemination of relevant content while maintaining the services' existing systems and processes.

3.229 Given the range of relevant content within scope of the measures (some of which may not be considered illegal or harmful in other jurisdictions), policies should set out sanctions for UK users to ensure the overall policy impact is proportionate.

3.230 Based on this evidence, and as explained in our June 2025 Consultation, we expect our measures to be proportionate and effective at reducing harmful relevant content across

---

[257] Ofcom (2024), Illegal Harms Register of Risks, chapter 25: Governance, systems and processes.
[258] A 2021 study found that posting warning messages to users believed to be at risk of suspension can reduce the use of hateful language on some social media networks. Yildirim, M.M., Nagler, J., Bonneau, R., and Joshua Tucker, A., 2023. Short of Suspension: How Suspension Warnings Can Reduce Hate Speech on Twitter, Perspectives on Politics, 21(2), pp.651-663.
[259] Electronic Arts (EA) state in its transparency report that they have found that warnings are their primary response to violations and that 80% of all violations are a first offence, suggesting that warnings reduce the risk of recidivism: EA, 2025. EA Player Safety Transparency Report 2024. Google LLC. said nine in ten people (94%) who receive a first strike never receive a second strike: Google LLC. response to 2022 Call for Evidence, p.30.
[260] TikTok said 90% of repeat violators on its platform violate the same feature consistently and that 75% violate the same policy category. TikTok, 2023. Supporting creators with an updated account enforcement system.
[261] Jhaver, S., Appling, D.S., Gilbert, E. and Bruckman, A., 2019. "Did you suspect the post would be removed?": Understanding user reactions to content removals on Reddit Proceedings of the ACM on human-computer interaction, 3, (CSCW); Suzor, N.P., West, S.M., Quodling, A. and York, J., 2019. What do we mean when we talk about transparency? Toward meaningful transparency in commercial content moderation International Journal of Communication, 13, p.18; Katsaros, M., Tyler, T., Kim, J. and Meares, T., 2022. Procedural justice and self-governance on Twitter: Unpacking the experience of rule breaking on Twitter. Journal of Online Trust and Safety, 1(3).
[262] For example, by de-ranking a user's profile or limiting access to their accounts for children if they have shared harmful or dangerous content: Northern Ireland Commissioner for Children and Young People (NICCY) response to the May 2024 Consultation on Protecting Children from Harms Online (May 2024 Consultation), p.33; NSPCC response to the May 2024 Consultation, pp.34, 47.

services. A range of stakeholders expressed support for the measures.[263] Many of these stakeholders noted the benefits of sanctions in affecting user behaviour, especially compared to content moderation alone.[264] Two stakeholders specifically agreed with the measure covering content harmful to children.[265] Three stakeholders specifically agreed with the application of the measures to illegal content.[266]

3.231 While stakeholder responses have not led us to fundamentally revise our assessment of the benefits of this measure, several stakeholders provided feedback or requested clarity on specific points of detail which could affect the efficacy of the measure. This included applying the measures to UK users, the importance of education and reform, feedback on the severity factor, consideration of intent as an additional factor, and requests for flexibility and/or the prescriptiveness of the measure. We discuss these responses in the following sub-sections and address additional stakeholder responses in the relevant chapter in annex 1: Further stakeholder responses.

Applying the measures to UK users

3.232 In response to our proposal to limit the measure to UK users, several stakeholders raised that this might lead to the measures being ineffective for certain harms.

a) Several fraud stakeholders said the illegal content measure should be extended beyond UK users to ensure it is effective at mitigating the harm.[267] We also received a number of requests from fraud stakeholders for the measure to include specific sanctions for fraudulent content.[268]

---

[263] Advertising Standards Authority (ASA) response to the June 2025 Additional Safety Measures Consultation, p.11; Association of British Insurers (ABI) response to the June 2025 Additional Safety Measures Consultation, p.3; Barnardo's response to the June 2025 Consultation, p.15; Center for Countering Digital Harm (CCDH) (new name) response to the June 2025 Additional Safety Measures Consultation, p.3. The Center for Countering Digital Harm, formerly The Center for Countering Digital Hate, changed its name since the June 2025 Consultation. We will refer to them by their new name for the remainder of the statement.; Check My Ads response to the June 2025 Consultation, p.11; Children's Commissioner's office (CCO) response to the June 2025 Additional Safety Measures Consultation, p.20; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.13; Evangelical Alliance response to the June 2025 Consultation, p.9; Financial Conduct Authority (FCA) response to the June 2025 Additional Safety Measures Consultation, p.5;[✂]; [✂]; ICO response to the June 2025 Consultation, p.24; Internet Matters response to the June 2025 Consultation, p.15; IWF response to the June 2025 Consultation, p.49.; [✂]; NCMEC response to the June 2025 Consultation, p.7; [✂]; National Police Chiefs' Council (NPCC) response to the June 2025 Additional Safety Measures Consultation, p.17; Refuge response to the June 2025 Additional Safety Measures Consultation, p.15;[✂]; SMACC response to the June 2025 Consultation, p.14; Which? response to the June 2025 Additional Safety Measures Consultation, p.13.

[264] Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.13; Evangelical Alliance response to the June 2025 Consultation, p.9; NCMEC response to the June 2025 Consultation, p.7.

[265] BILETA response to the June 2025 Consultation, p.28; Name Withheld 9 response to the June 2025 Additional Safety Measures Consultation, p.9.

[266] [✂]; [✂]; Which? response to the June 2025 Consultation, p. 13.

[267] ASA response to the June 2025 Consultation p.11; FCA response to the June 2025 Consultation, pp.5-6; [✂]; UK Finance response to the June 2025 Additional Safety Measures Consultation, p.13; Which? response to the June 2025 Consultation, p.15.

[268] ASA response to the June 2025 Consultation, p.11; Check My Ads response to the June 2025 Consultation, p.11; FCA response to the June 2025 Consultation, p.6; [✂]; UK Finance response to the June 2025 Consultation, pp.13-14; Which? response to the June 2025 Consultation, pp.14-15.

b) The Financial Conduct Authority (FCA) provided multiple points of feedback on how the measures applying only to UK users contrasts with the Financial Services and Markets Act 2000 (FSMA) and existing government priorities.[269]
c) Two respondents asked for clarity on how the measures work in the case of users using VPNs.[270]
d) Multiple stakeholders provided examples of how harms that originate from non-UK users can impact UK users.[271]

3.233 This is a difficult issue, and we recognise that users outside of the UK can post content and engage in conduct which causes significant harm to users in the UK. It is important that the regulatory regime protects UK users from such harm.

3.234 However, as noted in our June 2025 Consultation, there is a range of illegal content covered by the Act which will not all be illegal in other jurisdictions.[272] There is also a wide range of content harmful to children within the scope of the protection of children user sanctions measure; other jurisdictions may hold a different position on the severity of harm arising from such content.

3.235 In this context, we have concluded that, at this stage, it would not be appropriate to introduce a general measure on all harms which could result in overseas users being banned or otherwise sanctioned for content posted on a service which is not illegal or considered harmful in their jurisdiction. We consider that other measures and requirements of the Act will help to protect UK users from harmful content originating outside of the UK, including our proactive technology measures, the provisions around timely takedown of illegal content, and the measures around excluding potential primary priority content from recommender feeds of children to protect people in the UK from illegal content and content harmful to children which has been posted by overseas users.[273] We also note that providers have a duty under section 10 of the Act to take down illegal content to the extent that it can be encountered by UK users.

3.236 In the case of CSEA banning and proscribed organisations, we are satisfied that the balance of arguments is different because there is broad international consensus on the illegality of these harms and their severity. We have therefore taken a different approach in relation to these measures. We will keep under review whether it is appropriate to bring forward different bespoke sanctions measures for other individual harms in future.

3.237 We therefore consider our measures as consulted on are the most proportionate recommendation for addressing the risk of repeated harm and have decided to recommend providers prepare and apply a sanctions policy for UK users only.

3.238 We recognise that providers may choose to design their sanctions policy to apply globally. In practice, we expect providers will usually have a sanctions policy for non-UK users in any case, and if they wish, providers may incorporate our recommendations within their global policy that also provides for sanctions to be imposed on non-UK users.

---

[269] FCA response to the June 2025 Consultation, pp.6-7.
[270] BILETA response to the June 2025 Consultation, p.28; We Are Survivors response to the June 2025 Consultation, p.7.
[271] FCA response to the June 2025 Consultation, p.6; [✂]; Which? response to the June 2025 Consultation, pp.14-15.
[272] Paragraph 17.11, June 2025 Consultation.
[273] Children here means individuals deemed not to be adults by use of highly effective age assurance.

3.239 We have considered stakeholder feedback on the use of VPNs and evading sanctions and have addressed this in detail in the relevant chapter in annex 1: Further stakeholder responses.

Effectiveness of measure on services with existing sanctions policies

3.240 Some stakeholders called into question our assessment of the effectiveness and benefits of the measure. BILETA questioned the effectiveness of the measures where sanctions policies already exist.[274] The FCA said that Ofcom's non-exhaustive list of sanctions may lower standards compared to current practice.[275]

3.241 We are aware that many services already have sanctions policies, and some providers have replied to our June 2025 Consultation stating this.[276] In the consultation, we set out evidence that showed providers already use a range of sanctions.[277] However, there may be many services without a sanctions policy, and these measures are particularly important in setting expectations for those services.

3.242 Our measures do not just recommend providers have a sanctions policy. They also set expectations around consistency, clarity, and user redress, which we see as important to ensuring the measures are effective in mitigating harm and providing users clear guidance on expected behaviours on services.

3.243 Furthermore, we expect our measures will be beneficial because they cement good practice and make sure providers continue to have a sanctions policy that has regard to various relevant factors. From this, we consider they will ensure that all in-scope service providers have a sanctions policy, including those who may not now. In response to the FCA's comments, we therefore consider this measure does not lower standards but ensures all in-scope providers adopt effective sanctions policies.

3.244 Our measures reflect that policies vary between services and what is appropriate will depend on a service's attributes, such as size, type, and risk profile. As set out in the section 'How these measures work', providing this flexibility is important due to the number of services in scope of these measures, and the variety of risks that may arise on these services. Allowing providers to comply with the measures while determining their own sanctions policy provides the best means for them to achieve the intended outcome to reduce the future dissemination of relevant content.

3.245 Providers may choose to apply their existing sanctions policy where they are consistent with our measures, and we expect this would be effective in mitigating repeat offending.

3.246 We continue to consider that these measures as proposed would be effective in reducing the risk of harm to users and would provide benefits to users.

Impacts and costs

3.247 This section considers the direct and indirect impact of these measures on service providers, which we expect to vary considerably. As the measures are scalable and flexible,

---

[274] BILETA response to the June 2025 Consultation, p.29.
[275] FCA response to the June 2025 Consultation, p.5.
[276] Meta response to the June 2025 Consultation, p.12; [✂]; [✂]; [✂]; [✂].
[277] Google LLC. response to the 2022 Call for Evidence, p.29; Meta response to the 2022 Call for Evidence, pp.28-29; Reddit, 2025. Content Moderation, Enforcement, and Appeals; Signal, 2026. Signal Terms & Privacy Policy; Snap Inc., 2026. Snapchat Moderation, Enforcement and appeals; Twitch, 2025. About Account Enforcements and Chat Bans.

costs are likely to depend on factors such as the service size, level of risk, the type of sanctions, how the service provider chooses to implement them and existing systems and processes in place.

3.248 We consider costs are likely to increase with service size and risk. However, benefits are also likely to be larger due to the potential for larger reductions in the risk of harm on these services.

3.249 Costs are likely to be lower for service providers which already have a user sanctions policy in place. For service providers which need to include sanctions for both illegal content and content harmful to children in their policies, incremental costs may be smaller due to synergies between sanctions for the different types of content. Service providers within scope of these measures will also be within scope of the banning users for CSEA measure and there are likely to be some cost synergies with this, as outlined in paragraph 3.113.

3.250 In their responses to the June 2025 Consultation, BILETA, the Department for Justice for Northern Ireland, Digital Resilience in Education, Which? and two individuals said that they agreed with our assessment of the costs and impacts of these measures.[278] Which? noted that the cost to service providers is likely to be outweighed by the cost of doing nothing.[279]

Direct costs

3.251 In our June 2025 Consultation, we said that both one-off and ongoing costs would be incurred for policy development and communication of the policy to individuals working in content moderation.

3.252 There would also be one-off costs associated with changes to infrastructure and functionality, which would depend on what sanctions were proposed in the policy and how they were applied. There would also be ongoing costs associated with implementing the sanctions policy.

3.253 Due to the scalability and flexibility of these measures, we expected that all these types of costs would vary widely across service providers, increasing with service size and risk level.

3.254 In their responses to the June 2025 Consultation, one industry stakeholder and two individuals argued that these measures would place a disproportionate burden on smaller service providers that was not fully captured in our direct costs assessment.[280]

3.255 Ukie outlined that service providers, particularly smaller ones, were likely to incur broader costs when implementing the measures due to the need for additional resources to formalise, audit and report sanctions policy under the proposed measures.[281]

3.256 While we still consider that the costs of implementing a simple sanctions policy is likely to be low for the majority of small, low risk service providers, we acknowledge that our cost assessment did not fully reflect the costs to these providers of ensuring that they understand the specifics of the measures and ensuring they are confident that their sanctions policy meets the recommendations. Such service providers may require greater

[278] BILETA response to the June 2025 Consultation, p.30; Department for Justice for Northern Ireland response to the June 2025 Additional Safety Measures Consultation, p.14; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.14; [✂]; Name Withheld 9 response to the June 2025 Consultation, p.9; Which? response to the June 2025 Consultation, p.16.
[279] Which? response to the June 2025 Consultation, p.16.
[280] [✂]; Name withheld 11 response to the June 2025 Additional Safety Measures Consultation, p.2; [✂].
[281] Ukie response to the June 2025 Consultation, p.33.

policy development, specific legal advice, and review processes to be comfortable that their sanctions policy meets the criteria set out by the measures.

3.257 We therefore consider that the cost analysis for such providers included in our June 2025 Consultation is likely to under-estimate the true figure. As a result of this, we have also reconsidered our position on which service providers these measures should apply to (see 'Who these measures apply to' for further discussion).

3.258 Our assessment of the direct costs of these measures for larger and riskier service providers is unchanged from our position in the June 2025 Consultation. We expect that such providers may face higher costs as they will require a more complex policy and be required to implement it more frequently. However, we consider that these measures are proportionate for these providers, as explained in paragraphs 3.296-3.297. We have not provided a quantified cost estimate for such service providers, given that i) we are not prescribing specific actions and ii) services have the flexibility to implement the measure in a way that is appropriate.

Indirect effects

3.259 As explained in our June 2025 Consultation, these measures could also indirectly affect the ongoing costs of implementing existing measures, including staff time to deal with appeals and provision of training to staff working in content moderation. The measures may also have a potential impact on service user numbers and revenue. While it is difficult to assess these indirect impacts, we said that they were likely to increase with service size and risk.

3.260 A response from one industry stakeholder highlighted the impact of these measures on staff time for managing appeals and considered that this could divert resources from other safety or accessibility improvements.[282]

3.261 We acknowledge that these measures could indirectly result in more appeals, especially for high-risk services, resulting in more staff time and higher ongoing costs to implement the existing appeals measures. This impact has already been factored into our costs assessment. However, we also note that where user sanctions discourage or prevent future offending this could reduce costs.

3.262 Our assessment of the indirect impacts of these measures therefore remains unchanged from the assessment we made in our June 2025 Consultation, and we consider that any such effects will be proportionate.

Rights impact

Freedom of expression/Freedom of association

Our June 2025 Consultation

3.263 In the June 2025 Consultation, we recognised that users' rights to freedom of expression and to associate with others would be interfered with where a sanction limits their ability to interact with others on the service, such as a ban or a suspension. We also recognised that there may be an interference where lesser sanctions are applied and these deter users from sharing content, or where the prospect of sanctions deters them from doing so (in particular, children suffering from suicide ideation, self-harm, or an eating disorder). In our consultation we explained our view that the potential interference arising from these

[282] Ukie response to the June 2025 Consultation, p.33.

measures and any resulting sanctions would be justified and proportionate, noting that they pursue legitimate aims and promote the consistency and transparency of sanctions.

3.264 We also took into account the safeguards in ICU H2.7 and PCU H2.7 and PCU H2.8 for freedom of expression, including the existing appeals measures. We proposed to expand the scope of the appeals measures to allow them to function as appropriate safeguards for the user sanctions measures. We proposed amending the measures to include that, in determining appeals, providers should consider whether the sanction that has been applied is appropriate, having regard to its sanctions policy. We also proposed expanding the appeals measures so that where the provider upholds an appeal against a sanction, but finds the content still to be relevant content, they should reverse the action taken against the user and determine the appropriate sanction having regard to their sanctions policy.[283]

Feedback

3.265 In response to the June 2025 Consultation, Which? agreed with Ofcom's assessment of the impact on users' freedom of expression. Which? highlighted the significant public interest in reducing illegal content online and agreed that the interference with users' rights would be justified and proportionate. Which? also agreed that users' right to appeal against a sanction should ensure legitimate expression is protected, while illegal content is removed from services.[284]

3.266 Various stakeholders raised concerns that the measures have the potential to disproportionately interfere with legal speech, and therefore users' rights, [✂].[285] One stakeholder commented that sanctioning users for content proxy risked prohibiting lawful speech.[286]

3.267 Some stakeholders also raised concerns about service providers taking action against repeat offenders before criminal convictions have been secured, stating that it would be disproportionate to prevent a user from accessing a service due to their past actions.[287]

Interference

3.268 Users' rights to freedom of expression and to associate with others would be interfered with where a sanction is applied. To the extent that the objective of reducing harm compels a provider to adopt a policy committing it to impose sanctions, there is also an impact on providers' rights.

Prescribed by law

3.269 As mentioned in our June 2025 Consultation, an interference with the right to freedom of expression or freedom of association must be prescribed by, or in accordance with, the law.[288] Service providers of all user-to-user services have duties under section 10 of the Act to protect users from illegal content, and providers of services that prohibit content

---

[283] Illegal Content Codes, Measures ICU D8, D9, D10A; Protection of Children Code measures PCU D8, D9, D10A. For the full rights assessment in the June 2025 Additional Safety Measures Consultation, see paragraphs 17.75-17.90. We have since amended ICU D10A.2(b) and PCU D10A.2(b) to say that a provider should 'apply' the appropriate sanction, following determination of an appeal.
[284] Which? response to the June 2025 Consultation, p.16.
[285] [✂]; [✂]; Name Withheld 4 response to the June 2025 Consultation, p.8.
[286] [✂].
[287] Cook, J. response to the June 2025 Consultation, p.2; Name Withheld 4 response to the June 2025 Consultation, p.8.
[288] June 2025 Additional Safety Measures Consultation, paragraph 1.31, as set out in Articles 10(2) and 11(2) of the ECHR.

harmful to children have duties under section 12 of the Act to protect children from content that is harmful to them. Our measures recommend that providers have and apply a sanctions policy, with the aim of addressing their duties under sections 10 and 12 of the Act.

3.270 Although our measures give providers discretion as to what that policy contains, we are satisfied that any interference with users' rights will be prescribed by law, as our measure clearly sets out that a provider should have a sanctions policy, what that policy should achieve, and what factors a provider should take into account in setting the policy, those factors making clear where providers may consider more severe sanctions should apply. We did not receive any stakeholder feedback in response to the June 2025 Consultation regarding the interference with users' (or providers') rights being prescribed by law.

Necessary for a legitimate aim

3.271 The measure asks providers which are at risk of illegal and (where relevant) harmful content to prepare, apply and keep under review a sanctions policy with the objective of preventing future dissemination of such content by users who have disseminated it once. As such, it pursues the legitimate aims of the prevention of crime, the protection of health and morals, and the protection of the rights of others.

3.272 We consider that it is no more than is necessary to achieve that aim, because as set out in paragraph 3.226 there is evidence that offenders often engage in the same type of conduct online repeatedly, violating the same policy or same feature. Merely taking down the content concerned does not prevent reoffending. There is also evidence that better communication from services regarding the reasons for content removal and sanctions leads to increased perceptions of fairness and reduces likelihood of reoffending.

3.273 The measure builds in an ongoing evidence-based review of what works, ensuring that sanctions which are ineffective are not applied. We consider that providers have commercial incentives not to over-sanction users and are constrained by their duty under section 22 of the Act to have regard to impacts on freedom of expression.

Proportionate

3.274 Regarding concerns about the impact on rights to freedom of expression and association with others, the measures recommend that providers have regard to various factors when setting their sanctions policy. These include the severity of potential harm, and the potential impact of the sanction on the user being sanctioned (Measures ICU H2.4 and PCU H2.4). Having regard to these factors when setting their policy should lead to providers imposing sanctions that are appropriate and proportionate to the harm carried out.

3.275 As noted above, the measure does not prescribe the sanctions providers should include in their policy, as we expect providers will be best placed to decide on the appropriate sanctions to impose, having regard to the factors we have identified. We also note that providers will have reputational and commercial interests in maintaining users on their platform, which they will need to balance with protecting users from harm on their service and compliance with their duties under the Act. These factors should further ensure sanctions policies are set and enforced appropriately and proportionately.

3.276 The measures recommend that providers apply a sanctions policy in respect of the service on which a user has posted relevant content. As such, a user being sanctioned will only impact their rights to freedom of expression and association on that service. This limits the impact on user rights and ensures the measures are proportionate.

3.277 Measure PCU H2 applies only to services that are likely to be accessed by children, on which content harmful to children is prohibited. This means that there will only be interference with user rights in relation to lawful content where the provider has decided that such content is not permitted on its service.

3.278 As set out at paragraphs 3.328 to 3.330, we have amended measures ICU G1 and PCU G1 (now ICU G1A and PCU G1A as relating to user sanctions) to recommend that providers include information about their sanctions policy in their terms of service. Having a sanctions policy which is available to users should promote transparency and the consistent application of sanctions. This should help to ensure that users understand what sanctions they may face if they post illegal content or content that is harmful to children on the service.

3.279 Other measures also act as safeguards for users' rights, including the appeals and complaints measures, which have been amended as set out in paragraph 3.263.[289] These measures will be particularly relevant in the event of sanctions being imposed against users for content which has been incorrectly identified as being illegal content or content harmful to children, or where sanctions have been applied inconsistently with the service provider's policy, and will act as safeguards to users' rights. Measures ICU C4A and PCU C4A will also act as a safeguard where a provider reverses a decision that content is illegal content or illegal content proxy, other than in relation to a content appeal. For more detail on this measure, please see chapter 9: Deployment of proactive technology to detect image-based CSAM.

3.280 Regarding stakeholder comments on users being sanctioned for sharing proxy content, the measures recommend that providers sanction users with the objective of preventing the future dissemination of illegal content and content that is harmful to children. We recognise that a greater degree of interference with users' rights could arise if the service provider chooses to adopt terms of service which prohibit categories of content that are wider than required by the Act. We are not recommending under these measures that providers do this but recognise that in practice many providers will choose to apply their terms of service in order to comply with their duties. It is a matter of providers' own right to freedom of expression to decide what content they wish to allow on their service, as long as they protect UK users from the types of illegal or harmful content regulated by the Act. In practice, providers will have reputational and commercial interests which should incentivise them to apply an appropriate content proxy, and it will be their own decision to prohibit broader categories of content if they wish to do so.

3.281 Overall, we recognise our measures have the potential to interfere with users' rights to freedom of expression, and freedom of association. We recognise that this may occur even where a lesser sanction is applied, or where users are deterred from posting content due to the prospect of sanctions. This may particularly impact children, for example those who are suffering from suicidal ideation, self-harm, or an eating disorder, where providers of services which prohibit content harmful to children may impose sanctions for sharing such content. However, we consider that there are appropriate safeguards in place for any interference with users' rights and are therefore satisfied that any such interference arising

[289] Illegal Content Codes, Measures ICU D8, D9, D10A; Protection of Children Code, Measures PCU D8, D9, D10A.

from the measures is proportionate to the legitimate aims pursued. Similarly, we consider that the impact on providers' rights is proportionate.

Privacy

Our June 2025 Consultation

3.282 In the June 2025 Consultation, we acknowledged that enforcing a sanctions policy would have implications for users' rights to privacy but considered that any interference would be proportionate to the legitimate aims pursued by the measures.[290]

Feedback

3.283 In response to our consultation, one stakeholder agreed with our assessment of impact, while acknowledging the measures would impact users' right to privacy.[291] [✂].[292]

3.284 The ICO also provided comments regarding the privacy safeguards for the measure, suggesting that further elements of the draft codes should be included as privacy safeguards. Namely, the ICO considered that the amendments to measures ICU G1 and PCU G1, and elements of the content moderation measures would be likely to function as safeguards to users' right to privacy.[293]

Interference

3.285 Any interference with users' right to privacy will occur primarily where providers review content uploaded by the user which they suspect to be illegal, harmful to children, or proxy content, and process personal information for the purpose of implementing a sanction. We acknowledge that this will extend to enforcing sanctions, including preventing a user from returning to the service following the implementation of a ban (where that is determined to be the appropriate sanction by the provider).

In accordance with law

3.286 We explain in paragraphs 3.268 to 3.269 why our measures satisfy the requirement that any interference be 'prescribed by law'. For the same reasons we consider that the measures are 'in accordance with law' for the purpose of Article 8. We note that data protection law, in particular the principle of data minimisation, constrains providers' ability to process personal data without good reason.

Necessary for a legitimate aim

3.287 As set out in paragraphs 3.270 to 3.272, the measures pursue the legitimate aims of upholding the interests of public safety, preventing disorder or crime, and protecting health, morals, and the rights and freedoms of others. The evidence shows that users who upload illegal or harmful content are likely to do so repeatedly. Some interference with their privacy is the minimum necessary to take effective action to prevent or deter them from doing so, where such action is appropriate.

Proportionate

3.288 The harms the measure addresses are, in many cases, very serious. We have recommended that providers prepare their sanctions policy based on harm reduction and having regard to the seriousness of the harms concerned and the importance of users' rights to freedom of

---

[290] Ofcom (2025), Additional Safety Measures Consultation, paragraphs 17.88-17.90.
[291] Name Withheld 9 response to the June 2025 Consultation, pp.10-11.
[292] [✂].
[293] ICO response to the June 2025 Consultation, pp.24-25.

expression. That, together with providers' incentives not to over-sanction, means we consider the sanctions policy they arrive at will be proportionate. In implementing and enforcing their sanctions policy, we do not expect providers to process personal data beyond that which is necessary, nor retain personal data for longer than needed. We consider that the interference with users' privacy is therefore likely to be limited.

3.289 As set out in chapter 1: Introduction, our duties and navigating the Statement, we acknowledge the importance of the measures highlighted by the ICO to help providers comply with data protection principles, and note that providers will also remain subject to applicable data protection principles.

3.290 Other measures also provide safeguards for users' right to privacy. Providers must have suitable reporting and complaints processes in place including in relation to appeals. Where appropriate, providers must overturn or change a sanctions determination against a user. This is particularly important where a provider has incorrectly applied a sanction, for example where they have incorrectly identified illegal content.

3.291 With regard to the safeguards for privacy for the user sanctions measure, we consider that the safeguards as currently set out encompass all those relevant to users' right to privacy.

3.292 Overall, we consider that any interference with users' right to privacy will be justified and proportionate.

Who these measures apply to

3.293 In our June 2025 Consultation, we proposed that measure ICU H2 should apply to all user-to-user service providers, and measure PCU H2 should apply to all user-to-user service providers, likely to be accessed by children, which prohibit at least one specific type of primary priority content, priority content, and/or non-designated content.

3.294 As set out in the section 'Impacts and costs', we consider that our cost estimates in the June 2025 Consultation may not have fully reflected the costs to small, low risk service providers. We also note that the benefits to users of such service providers having a formal sanctions policy would be lower, given that the smaller number of users means that such a policy might only be applied infrequently.

3.295 Given the above, we have decided that these measures should not apply to small providers of services with a low level of risk of relevant content, or which are at medium to high risk of only one type of harmful content (which we refer to as 'single risk'). They do however apply to providers of smaller services where they are multi-risk for relevant content.[294]

3.296 Where such service providers already apply sanctions to users who share illegal content or content harmful to children, they can continue to do so. However, they should be able to choose to do this on a case-by-case basis rather than having to develop a formal policy in line with the measures.

3.297 Although the costs of these measures are likely to be higher for service providers which are multi-risk for relevant content, we consider that the benefits will outweigh the costs due to the higher volume of relevant content present on such services, even where they are small. The measures give service providers the flexibility to create a sanctions policy that will best

[294] Small services are those with an average user base of fewer than seven million monthly active UK users. This is consistent with the approach taken in the Illegal Content and Protection of Children Codes, where 'large services' are defined as those with more than seven million monthly active UK users.

work for their service. We therefore consider that it is proportionate to apply these measures to multi-risk service providers of all sizes.

3.298 Applying these measures to large service providers, even where these are low risk, is likely to have significant benefits for users. The size of such services means that the presence of harmful content has the potential to affect many users, and the nature and prevalence such content can change over time, meaning that even if a large service is currently low risk, this could change quickly. Having a sanctions policy in place will help to ensure that if there were to be an increased risk of harm on such services, this would be dealt with quickly, reducing harms which on a large service would have the potential to affect many users. Larger service providers are also likely to have sufficient capacity and resources to be able to implement these measures. We therefore consider that it is proportionate to apply these measures to large service providers regardless of risk level.

3.299 We are therefore applying the following scope for these measures:

a) Measure ICU H2 will apply to providers of user-to-user services that are large and/or multi-risk in relation to illegal content; and
b) Measure PCU H2 will apply to providers of user-to-user services that are likely to be accessed by children, prohibit at least one specific type of primary priority content, priority content, and/or non-designated, and are large and/or multi-risk in relation to content harmful to children.

3.300 These measures do not apply to search service providers, as users of these services do not generate, upload, or share content.

Conclusion

3.301 Our analysis shows that services having a sanctions policy will provide benefits to UK users. By ensuring that providers are working towards the objective of preventing future dissemination of relevant content, we expect a reduction in repeat sharing of relevant content.

3.302 Our measures offer providers an appropriate balance of flexibility to implement the measures in a way that works best for their services, while setting out factors to which they should have regard and practice to ensure sanctions are effective and applied consistently.

3.303 We consider that the benefits of the measures are more substantial on services which are either large or multi-risk, and that the costs are justified. We recognise that it may not be proportionate for smaller services with low/single risk to be in scope of the measures when weighing up the benefits of the measures to users and have therefore decided not to apply the measures to them.

3.304 Many stakeholders agree with the measures, and industry stakeholders have provided evidence of existing sanctions policies. Having assessed the remaining stakeholder responses, and based on our analysis of them, we have decided to confirm these measures. We consider it proportionate to apply these measures to providers of those services outlined in paragraph 3.298.

3.305 We are including these measures in our Illegal Content and Protection of Children Codes of Practice, where they are referred to as ICU H2 and PCU H2. The measures are part of our Codes of Practice for user-to-user services on Terrorism, Other duties, and Protection of Children.

Amendments to the Codes to reflect user sanctions policy measures and user banning measure

3.306 As a result of the user sanctions policy measures and the CSEA banning measure proposed in our June 2025 Consultation, we also proposed changes to several existing measures.[295]

3.307 We proposed changing our existing appeals measures in relation to the CSEA banning measure as follows:

a) Where the provider does not reverse a decision that content was CSEA or CSEA content proxy, but upholds, in whole or in part, a content appeal relating to the user being banned, the provider may reverse the actions taken where exceptional circumstances apply or change the duration of the ban.[296] This amendment is discussed in detail in paragraphs 3.321 to 3.325 of this section.

i) We set out the extenuating circumstances in which a provider may reverse a ban against a user whom they have reasonable grounds to infer is a child. A ban may only be reversed in two circumstances against a child: where the user has shared, generated or uploaded SGII as a result of having been groomed; or where they have consensually shared SGII and/or received SGII consensually as part of an age-appropriate relationship.[297] Our decision to confirm this amendment is discussed in detail in the CSEA banning section from paragraph 3.85, and is not covered further in this section.

3.308 We proposed changing our existing appeals measures in relation to the user sanctions measures as follows:

a) In determining appeals, providers should consider whether the sanction that has been applied is appropriate, having regard to its sanctions policy.[298]
b) Where the provider upholds an appeal against a sanction (in whole or part) but finds the content still to be illegal or harmful content (or proxy content), the provider should reverse the action taken against the user and determine the appropriate sanction, having regard to its sanctions policy.[299]

3.309 In relation to both user sanctions and user banning, we proposed changing our existing terms of service measures so that providers include provisions giving information about their policy on sanctions in their terms of service.[300]

3.310 And finally, in relation to user sanctions, we proposed changing our existing content moderation measures on performance targets to recommend that certain providers should set performance targets for the time period for reviewing and assessing content. These

[295] For CSEA banning, see paragraphs 16.80-16.86 of our June 2025 Consultation. For user sanction policies, see paragraphs 17.45-17.48 of our June 2025 Consultation for drafting on appeals and paragraph 17.49 for drafting on terms of service.
[296] Illegal Content Codes, Measure ICU D10.4 (as in the June 2025 Consultation, now ICU D10.3).
[297] Illegal Content Codes, Measure ICU D10.5. (as in the June 2025 Consultation, now ICU D10.3).
[298] Illegal Content Codes, Measure ICU H2.6 (referring to ICU D8 and ICU D9) (as in the June 2025 Consultation, now ICU D8.5); Protection of Children Code, Measure PCU H2.6 (referring to PCU D8 and PCU D9) (as in the June 2025 Consultation, now PCU D8.5).
[299] Illegal Content Codes, Measure ICU D10; Protection of Children Code, Measure PCU D10 (as in the June 2025 Consultation, now ICU D10A and PCU D10A respectively).
[300] Illegal Content Codes, Measure ICU G1; Protection of Children Code, Measure PCU G1 (as in the June 2025 Consultation, now ICU G1A and PCU G1A respectively).

providers were those to whom part of the performance targets measure regarding targets for relevant content moderation action did not previously apply outside of CSEA and proscribed organisation content.[301]

3.311 See chapter 7: Broadening content appeals for amendments to reporting and complaints measures on content proxies.

Our decision

3.312 We have decided to broadly confirm these amendments as proposed in the June 2025 Consultation and outline our reasons for doing so in the next section.

3.313 In respect of the banning users for CSEA measure, in response to stakeholder feedback, and in line with amendments we have made in respect of the treatment of children, we have decided to amend appeals measure ICU D10.3. We have set out that on appeal, where the provider has reasonable grounds to believe the user is a child and where:

a) one of the two extenuating circumstances exists, they may choose to overturn the ban;
b) an extenuating circumstance does not exist, they may amend the duration of the ban, taking into account the factors set out at paragraph 3.94.

3.314 In response to stakeholder feedback, and in line with the changes we have made regarding how users who receive and view CSAM should be treated under this measure, we have decided to amend appeal measure ICU D10.3. This measure now says that, on appeal, where the provider is satisfied that the user promptly reported and deleted the content, the provider may choose to overturn the ban.

3.315 We have also made some other minor amendments to measure ICU H3. These amendments are not intended to reflect material changes to the measures but have been made to ensure clarity and consistency in our Illegal Content Codes. The amendments are as follows:

a) The definition of "age-appropriate relationship" is now set out in the body of the code measure, at ICU H3.2(a). Previously we had proposed to include this definition in the definitions Table A of the codes.
b) We have separated the definition of a "relevant user" in ICU H3.2(c) from the recommendation of which users should be banned under the measure, at ICU H3.3 and ICU H3.4. Previously this recommendation and definition were both set out in ICU H3.2.
c) We have separated the reasonable steps providers should take to prevent a banned user from regaining access to the service, making clear the provider should have regard to these factors as set out in ICU H3.7 (ICU H3.4, as recommended in the June 2025 Consultation).
d) We have amended ICU H3 to refer to a provider having "reasonable grounds to believe" a user is a child, as opposed to "reasonable grounds to infer" (as recommended in the June 2025 Consultation).

[301] The amendment proposed to extend measures to providers of services that are large and/or multi-risk for illegal harms, for whom it is currently not technically feasible to take down content, and providers of services (i) that are large and/or multi-risk for content harmful to children; (ii) prohibit one or more specific kinds of content harmful to children; and (iii) for whom it is currently not technically feasible to take content level action on content harmful to children, or technically feasible (or, in the case of priority content, proportionate) to apply access controls.

e) We have included additional safeguards at ICU H3.12 (ICU H3.7 as recommended in the June 2025 Consultation)

3.316 Where we have amended our user sanctions measures, apart from our performance targets measure, we have also made a change to who they apply to, removing providers of small services that are single or low risk from the scope of amendments relevant to these measures.[302] See the section 'Who these amendments apply to' in relation to the user sanctions measure for our full reasoning. As a result of this amendment, the appeals measures and terms of service measures relating to user sanctions have now been separated into separate measures, namely ICU D10A, PCU D10A, ICU G1A and PCU G1A.

3.317 We have also made some other minor amendments to measure ICU H2. These amendments are not intended to reflect material changes to the measures but have been made to ensure clarity and consistency in our Illegal Content and Protection of Children Codes. The amendments are as follows:

a) We have amended measures ICU H2.6 and PCU H2.6 to move these provisions to our measures on reporting and complaints. These measures can now be seen at ICU D8.5 and PCU D8.5 respectively.
b) We have amended ICU H2.3 to remove the cross-reference to determinations made under ICU C1 and ICU C2 regarding determining illegal content or illegal content proxy.
c) Further safeguards for freedom of expression and privacy have been added at ICU H2.6.

3.318 The full text of the measures can be found in the Illegal Content and Protection of Children Codes, and are referred to as:

a) ICU D8, PCU D8,[303] and ICU D10, ICU D10A and PCU D10, PCU D10A for the amended appeals measures;[304]
b) ICU G1, ICU G1A, and PCU G1, PCU G1A for the amended terms of service measures; and
c) ICU C4 and PCU C4 for the amended content moderation measures.

3.319 The measures are part of our Illegal Content Codes on CSEA, Terrorism and Other duties and in our Protection of Children Code.

How these amendments work

Appeals

3.320 Providers have a duty to operate a complaints procedure that allows for relevant complaints to be made and provides for appropriate action to be taken by the provider in response to such complaints.[305]

---

[302] This maintains our position for Measures ICU C4 and PCU C4, which have always only applied to large and/or multi-risk services.
[303] Measures ICU D8 and PCU D8 now include the recommendation that, in determining appeals, providers should consider whether the sanction that has been applied is appropriate in having regard to its sanctions policy. In the June 2025 Consultation, this was found in ICU H2.6 and PCU H2.6. For our final measures, we have moved this to ICU D8.5 and PCU D8.5.
[304] Note that with changes to whom these user sanctions measures apply, Measures ICU D9 and PCU D9 are no longer captured in this amendment.
[305] Providers have a duty under Section 21 of the Act to operate a complaints procedure that allows for relevant complaints to be made and provides for appropriate action to be taken by the provider in response to such complaints.

3.321 Our Illegal Content and Protection of Children Codes recommend that providers should determine relevant complaints which are content appeals.[306] Appeals include complaints by a UK user if the user has been given a warning, suspended, banned, or in any other way restricted from using a service, as a result of content generated, uploaded or shared by the user which the provider considers to be illegal or content harmful to children.[307]

CSEA banning

3.322 Our amendment to the appeals measure (as described in the 'Appeals' section) provides an important safeguard and protects users in cases where a ban might be disproportionate or overly punitive.

3.323 We have also decided to amend the appeals measure ICU D10.3, in accordance with our amendments to the user banning measure in respect of the treatment of children. The appeals measure will allow providers to re-determine the appropriate duration of a ban where they did not have reasonable grounds to believe that the user was a child when originally deciding to ban but are then presented with evidence that gives them reasonable grounds to believe the user was a child. The measure will also allow providers to re-determine the appropriate duration of a ban where they did have reasonable grounds to believe the user was a child when applying the ban, but the child subject to the ban has appealed on the basis that the duration of the ban was inappropriate. In these cases, the provider should consider the appropriate duration of the ban or, if an extenuating circumstance exists, may decide to overturn the ban.

3.324 As set out above, we have also decided to amend the appeals measure ICU D10.3, in accordance with our amendments to measures ICU H3.4 and ICU H3.5 regarding the banning of users who have received and viewed CSAM. We recognise that providers may not have information as to whether a user has promptly reported and deleted the content at the time of applying a ban, and users may wish to submit an appeal against the ban to provide this information. This amendment to measure ICU D10.3 will allow providers the option to overturn a ban on appeal where they are satisfied the user promptly reported and deleted the content.

3.325 While the ICO said that it supports our changes to broaden the scope of appeals to enable users to appeal decisions relating to illegal content proxy, as this is an important privacy safeguard, it also asked for clarity on how CSEA content proxy may be treated at appeal.[308] We acknowledge that the proposed amendments to the reporting and complaints measures would not allow for providers to overturn a ban where they determine that content proxy is legal but nevertheless still in breach of their terms of service (in which case the content would still be content proxy for the purpose of our Codes). This includes where a provider prohibits all nudity content in its terms of service and uses this as a proxy to follow this measure. In deciding whether to follow this measure by applying a proxy, and if so the nature of that proxy, we expect providers will need to balance various commercial and reputational factors such as the cost of compliance and user retention.

3.326 Providers are also required to comply with data protection law, under which they must be transparent and fair about how they process users' data, including the circumstances in

[306] Illegal Content Codes, Measures ICU D8; Protection of Children Code, Measure PCU D8.
[307] In chapter 7: Broadening content appeals we propose amendments to ensure that the appeals measures cover decisions taken on the basis that content was illegal or content harmful to children proxy.
[308] ICO response to the June 2025 Consultation, p.22.

which they will ban a user. We expect providers employing a proxy to make clear to users the content for which they will be banned, and that the provider will not overturn a ban where it finds content was not illegal CSEA but was nevertheless in breach of terms of service. See the 'Rights impact' section of this chapter for an assessment of the impact on user rights of including CSEA content proxy, and paragraph 3.129 for a discussion of the costs.

User sanctions

3.327 We have now made amendments to the appeals measures to recommend that providers should consider whether, in the circumstances of the individual case, the sanction that has been applied is appropriate having regard to its sanctions policy. We have amended the wording of ICU D10A.2 and PCU D10A.2 (ICU D10.3 and PCU D10.3 as recommended in the July 2025 Consultation) to refer to a user having been sanctioned under the provider's policy on sanctions, as opposed to being sanctioned *pursuant to* the provider's policy on sanctions. This amendment has been made to ensure users can still appeal against a decision which was not imposed consistently with the provider's policy on sanctions.

3.328 We are also recommending that if the provider does not reverse a decision that content was relevant content but upholds (in whole or part) a content appeal relating to a user sanction, the provider should reverse the action taken against the user and apply the appropriate sanction having regard to its sanctions policy.[309]

Terms of service

3.329 Providers have a duty to include provisions in their terms of service specifying how individuals are to be protected from illegal content, how children are protected from content harmful to them, any proactive technology used and provide users with information on how complaints are handled.[310]

3.330 Our measures now state that providers should include information about the sanctions policy in their terms of service to provide transparency to users.[311] This should include at least the types of sanctions that the provider may impose under its policy.

3.331 While we are making amendments to existing terms of service measures regarding details about sanctions, the measure does not provide that the full detailed policy should be included in external publications.

Performance targets

3.332 Measures ICU C4 and PCU C4 in our Illegal Content and Protection of Children Codes previously stated that providers of large and/or multi-risk services should set and record performance targets covering at least the time period for taking relevant content moderation action and the accuracy of decision making.[312]

3.333 In our June 2025 consultation, we proposed to amend these measures as follows:

---

[309] Illegal Content Codes, Measure ICU D10.3; Protection of Children Code, Measure PCU D10.3. At consultation, we said providers should 'determine' the appropriate sanction. For our final measures, we have changed this to recommending that providers 'apply' the appropriate sanction, following determination of the appeal under Measures ICU D8.5 and PCU D8.5.
[310] Section 10(8).
[311] Illegal Content Codes, Measure ICU G1A; Protection of Children Code, Measure PCU G1A.
[312] Relevant content moderation action includes reviewing and any content level action.

a) Measure ICU C4 previously recommended that providers of services on which it is currently not technically feasible to take down content take the action recommended by ICU C1.3 (making an illegal content judgement or determining whether content is in breach of terms of service) in respect of at least suspected CSEA content and suspected proscribed terrorist organisation content. [313] We proposed to amend the measure to recommend that providers take the action in ICU C1.3 in respect of all suspected illegal content.
b) We proposed to amend measure PCU C4 to recommend that, for providers of services on which (i) it is currently not technically feasible to take content level action and it is not technically feasible (or, in the case of priority content, proportionate) to apply access controls and (ii) one or more specific kinds of content that is harmful to children are prohibited, they take the action recommended in PCU C1.3 to PCU C1.5 (consider whether content is in breach of its terms of service or otherwise content harmful to children), in so far as it relates to the kinds of content that are prohibited.

3.334 The reason for these proposed amendments was because, in the user sanctions policy measures, we proposed that providers can take action against users who post relevant content, even where it is not technically feasible to take action on the content itself.

3.335 As we are proceeding with the user sanctions policy measures, we have decided to confirm that providers of services on which it is not currently technically feasible to take down content should, as set out above, set performance targets to take relevant content moderation action in relation to relevant content. Providers should then apply their sanctions policy to users who have posted such content (which may mean deciding not to impose any sanction). We have decided to specify that providers should set performance targets for taking action in relation to the kinds of relevant content for which a user may be sanctioned under their policy. This is because it is in respect of these types of content that timely content moderation action and application of the sanctions policy may protect users from harm.

Our reasoning

Effectiveness and benefits, and costs impact

3.336 As stated in our June 2025 Consultation, we consider our amendments will help reduce harm to UK users in several ways. In regard to performance targets, we consider striking a balance between timeliness and accuracy is important. The faster such providers review and assess relevant content, the faster they will be able to take appropriate action under their user sanctions policies. As set out 'How these amendments work' changes to appeals and terms of service measures will bring improved transparency and empowerment to users. [314]

[313] Providers can report CSEA to NCMEC or the NCA and remove accounts run by proscribed organisations. See the December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement), Volume 2, paragraph 2.231.
[314] For the full rights assessment regarding amendments to the terms of service and appeals measures, see the June 2025 Consultation, paragraphs 17.54-17.90. For the rights assessment regarding the amendments to performance targets measures, see paragraphs 17.103-17.107.

Performance targets

3.337 In relation to our performance targets amendment, one stakeholder shared its own practice in regard to setting performance targets for its moderation systems.[315] Another provided evidence highlighting the importance of performance targets in motivating service providers to swiftly sanction users.[316]

3.338 As addressed in previous statements, we are aware of providers setting other targets for their moderation systems, and our measures do not preclude providers from continuing to use these.

3.339 Regarding performance targets, we consider that the costs assessment for the original measure which we consulted on remains appropriate for the purposes of considering this amendment.[317]

3.340 The costs assessments in these documents related to the setting of performance targets for content moderation systems related to both speed and accuracy.[318] We said that the flexibility given to service providers regarding how to implement these measures means that costs are likely to vary widely between providers. The upfront costs to a smaller service provider of designing or selecting a small and relatively simple set of metrics would be limited.[319] However, for larger service providers or those with many risks, these costs could be more significant – potentially up to tens or hundreds of thousands of pounds, due to the number and complexity of metrics involved and the associated data management processes. There would also be ongoing costs to measure performance against these metrics and analysing these metrics as well as data storage costs, but again these costs are likely to vary significantly between service providers.[320]

3.341 We note that service providers in scope of the performance target amendment will already have in place performance targets relating to accuracy, and some may have time targets relating to CSEA and proscribed organisation content. Therefore, there may be some cost synergies with existing performance targets.

Appeals and terms of service

3.342 Regarding appeals, we consider that service providers are likely to use existing systems and processes to deal with these amendments, meaning that additional associated costs are likely to be small. Likewise, we consider that the cost to providers of adding information about their user sanctions policy to their terms of service is likely to be minimal.

---

[315] Snap Inc. response to the June 2025 Consultation, p.14.
[316] FCA response to the June 2025 Consultation, p.10.
[317] We consulted on these costs in our November 2023 Consultation and our May 2024 Consultation. They were included in our December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement) and our April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement).
[318] Illegal Content Codes, Measure ICU C4; Protection of Children Code, Measure PCU C4.
[319] Potentially in the region of £9,000 to £19,000 based on an assumption of 30 days of software engineering time being required. Note that these figures are slightly different to those referenced in earlier publications as they have been updated to reflect latest ONS data. Alternatively, a service provider might opt to licence a third-party ticketing system at a relatively low cost. Such solutions are available from around £50 per month per staff user.
[320] For full details of our previous cost assessments, see our November 2023 Consultation, paragraphs 12.108-12.110, December 2024 Statement paragraphs 2.235-2.238, May 2024 Consultation paragraphs 16.124-16.130, and April 2025 Statement paragraphs 14.253-14.256.

3.343 These amendments would form part of a package of measures relating to content moderation for illegal content and content harmful to children. As explained in our December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement) and in our April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement), and taking into consideration the benefits to users and other affected persons, we consider that any impact of measures ICU C4 and PCU C4 on rights to freedom of expression and association, and on privacy and data protection rights, is proportionate. This is because risks raised by the measure regarding speed targets and excessive takedown are mitigated by the inclusion of accuracy targets, and the flexibility of the measure in recommending providers strike a balance between this a time targets and can set other targets.[321]

3.344 As set out in our rights assessment in the user sanctions measure, the amendments to the existing appeals and terms of service measures are safeguards and beneficial for users' rights in relation to sanctions policies. We also explain in our rights assessment for the CSEA banning measure that the amendments to the appeals measures are safeguards for user rights.

3.345 As we have noted the effectiveness and flexibility of the measures, and our understanding of the costs and the safeguards in place for users' rights, we do not see any reason to amend the measures as proposed in our consultation.

3.346 Stakeholders were broadly supportive of these amendments. One highlighted the need for flexibility in setting performance targets (see relevant chapter in annex 1: Further Stakeholder responses for feedback).

Who these measures apply to

Performance targets

3.347 In our June 2025 Consultation, we proposed that providers of the following services should set performance targets for the time period for reviewing and assessing content, and track whether they are meeting these:

a) services that are large and/or multi-risk for illegal harms, for whom it is not currently technically feasible to take down content; and
b) services that are large and/or multi-risk for content harmful to children and prohibit one or more specific kinds of content harmful to children, for whom it is not currently technically feasible (or, in the case of priority content, proportionate) to apply access controls.

3.348 In response to our proposed changes to performance targets measures in the June 2025 Consultation, Middle Tech Coalition expressed concern about the proposed additional performance targets for content moderation, stating that their concern was for the extension of performance targets measures to services where it is not technically feasible to take down content.[322]

3.349 We do not think our proposed changes present any interference with services where takedown is not technically feasible. As with suspected CSEA and proscribed organisation content, our user sanctions measures now provide an action for providers, where it is not

[321] Ofcom (2024), December 2024 Statement, Vol.2, chapter 2: Content moderation; Ofcom (2025), April 2025 Statement, Vol.4, chapter 14: Content moderation for user-to-user services.
[322] Middle Tech Coalition response to the June 2025 Consultation, p.6.

technically feasible for content takedown, to set performance targets for the time period of reviewing and assessing content. This serves to maintain these providers' existing moderation practices, while including targets to be held accountable to. As such, they should not create a material additional burden. We therefore conclude this change is justified.

3.350 We have accordingly confirmed our decision to apply the performance target change to the service providers outlined in paragraph 3.346.

Appeals and terms of service

3.351 In our June 2025 Consultation, we proposed that the changes to appeals and terms of service measures should apply to all user-to-user service providers.

3.352 However, as explained in paragraph 3.293, we do not now consider that the user sanctions measures are proportionate for providers of smaller services which are low/single risk, and recommend that these measures apply only to user-to-user service providers that are large and/or multi-risk in relation to the relevant harms. Given that we are not recommending that smaller, lower/single risk service providers will have a user sanctions policy, we do not consider that it is logical or proportionate for them to be included in changes to other measures which support the user sanctions measures.

3.353 We have therefore decided not to proceed with the proposed amendments to measures ICU D9 and PCU D9, as these apply only to services which are small and low/single risk.

3.354 We are recommending that:

a) the changes to measures ICU D8, ICU D10A and ICU G1 (now ICU G1A) in relation to the user sanctions policy apply to providers of user-to-user services that are large and/or multi-risk for illegal harms; and
b) the changes to measures PCU D8, PCU D10A and PCU G1 (now PCU G1A) in relation to the user sanctions policy will apply to providers of services that are large and/or multi-risk for content harmful to children.

3.355 We confirm our proposal in the June 2025 Consultation that the changes to appeals measures regarding CSEA should apply to all user-to-user service providers, to be consistent with our decision on the user banning measure.

Conclusion

3.356 Our analysis shows that these amendments offer important safeguards for users that ensure that users can appeal and understand a service's sanctions policy, and that providers are taking steps to set targets and monitor the effectiveness of the moderation systems that underpin these policies. We conclude that the changes to these measures are appropriate and proportionate.

3.357 As set out in paragraph 3.293, for small, low/single risk services, it would not make sense to apply these changes given that we are not including such providers within the scope of the user sanctions policy measure.

3.358 The measures are part of our Illegal Content Codes on CSEA, Terrorism and Other duties and in our Protection of Children Code.

4. User controls

What is this chapter about?

User controls include a range of functions to allow a user to control their interactions on online services, including functions such as user blocking, muting or disabling comments. In our April 2025 User Controls Consultation, we proposed expanding existing user control measures to include providers of certain smaller user-to-user services that are likely to be accessed by children, have relevant functionalities and are at risk of certain illegal harms.

The harms we seek to address include controlling or coercive behaviour, cyberflashing, harassment, stalking, threats and abuse, hate, grooming and encouraging or assisting suicide or serious self-harm.

What decisions have we made?

We have decided to proceed with our proposed extension. Our final Code measures are summarised in the table below:

Number in Codes	Recommended measure	Who should implement this
ICU J1	Service providers should offer every registered user the options to block and mute other user accounts on the service.	Providers of user-to-user services likely to be accessed by children that have user profiles, have at least one of user connections, posting content, and user communication, and have: i) no more than seven million monthly UK users and high risk for at least one of grooming; encouraging or assisting suicide or serious self-harm; hate; harassment, stalking, threats and abuse; controlling or coercive behaviour; cyberflashing; or ii) between 700,000 and seven million monthly UK users and medium risk for at least one of the relevant harms.
ICU J2	Providers of large user-to-user services should offer every registered user the option to disable comments on their content.	Providers of user-to-user services likely to be accessed by children that have the functionality of commenting on content, and have: i) no more than seven million monthly UK users and high risk for at least one of grooming; encouraging or assisting suicide or serious-self harm; hate; harassment, stalking, threats and abuse; or ii) between 700,000 and seven million monthly UK users and medium risk for at least one of the relevant harms.

Why have we made these decisions?

Functionalities including direct messaging, posting and commenting on content can be a risk factor for experiencing harms online. We know that user controls such as blocking, muting, and disabling comments are widely used by users to protect themselves from harm.

The extension of these measures to small services will significantly reduce the risk of child users being exposed to these illegal harms. Children will also be provided with clear information to help them understand what will happen if they implement these controls. This will empower children to control the content they see on services, who can contact them and who can interact with them, ensuring that children receive a higher standard of protection as required by the Act.

Introduction

4.1 Alongside the April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement), we consulted on expanding the application of our user blocking and muting measure and the disabling comments measure in the Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes).[323]

4.2 In this chapter, we address the feedback we received on expanding these measures and set out our final decision. As stakeholders commented on the measures together, we have responded to this feedback in a similar manner, drawing out any comments on individual measures where appropriate.

Measures on user blocking and muting, and disabling comments

4.3 In the December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement) and the Illegal Content Codes, we recommended user control measures for large user-to-user service providers with relevant functionalities that identified the risk of certain illegal harms. These included:

a) User blocking and muting (ICU J1): Providers of large user-to-user services should offer every registered user the options to block and mute other user accounts on the service, if they have:

   i) identified as medium or high risk for any of coercive or controlling behaviour; harassment, stalking, threats and abuse; hate; grooming; encouraging or assisting suicide;
   ii) user profiles; and
   iii) at least one of the following functionalities: user connections; posting content; user communication (including but not limited to direct messaging and commenting on content).

b) Disabling comments (ICU J2): Providers of large user-to-user services should offer every registered user the option of disabling comments on their content, if:

[323] Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes), Measures ICU J1 and ICU J2.

i) they have identified as medium or high risk for any of harassment, stalking, threats and abuse; hate; grooming; or encouraging or assisting suicide; and
ii) users are able to comment on content.

4.4 In our April 2025 User Controls Consultation (April 2025 Consultation), we consulted on expanding the application of these measures in our Illegal Content Codes.[324] This expansion proposed to bring providers of certain smaller user-to-user services that are likely to be accessed by children into scope of the measures where they have relevant risks and functionalities.

4.5 We said that providers should ensure that users who have not been determined to be adults are offered these controls. Providers could do this by using highly effective age assurance to distinguish adult users from other users, or, where they do not use highly effective age assurance to do this, they could instead offer the measures to all users on the parts of the service that are accessible by children.

4.6 We explained that the measures would still apply for all users in the case of providers of large user-to-user services with the relevant risks and functionalities, as currently set out in the Illegal Content Codes.

4.7 We proposed this extension to reduce the risk of children encountering the relevant harms. We set out our view that this was justified to reflect the higher standard of protection for child users envisaged by the Online Safety Act 2023 (the Act).

4.8 In summary, we proposed the following:

a) User blocking and muting: The measure would apply to providers of user-to-user services likely to be accessed by children that have user profiles, have at least one of the following functionalities (user connections, posting content, user communication), and have:

i) below seven million monthly UK users and high risk for at least one of coercive or controlling behaviour; harassment, stalking, threats and abuse; hate; grooming; encouraging or assisting suicide; or
ii) between 700,000 and seven million monthly UK users and medium risk for at least one of the relevant harms.

b) Disabling comments: The measure would apply to providers of user-to-user services likely to be accessed by children that have the functionality of commenting on content, and have:

i) below seven million monthly UK users and high risk for at least one of harassment, stalking, threats and abuse; hate; grooming; encouraging or assisting suicide; or
ii) between 700,000 and seven million monthly UK users and medium risk for at least one of the relevant harms.

[324] In the December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement) and the Illegal Content Codes, we recommended user control measures for large user-to-user service providers with relevant functionalities that identified the risk of certain illegal harms. In our April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement), we also recommended these measures in our Protection of Children Code of Practice for user-to-user services (Protection of Children Code) to services likely to be accessed by children, including certain smaller services with relevant functionalities where they identified the risks of certain kinds of content harmful to children.

Our decision

4.9 We have decided to proceed with the extension of these measures as proposed in our April 2025 Consultation. The full text of the measures can be found in our Illegal Content Codes and are referred to as ICU J1 and ICU J2.[325] The measures are part of our Illegal Content Codes of Practice for user-to-user services on child sexual exploitation and abuse (CSEA) and other duties.

4.10 As explained in chapter 1: Introduction, our duties and navigating the Statement, two new priority offences have been introduced into the Act. Due to this, the blocking and muting measure now also applies to serious self-harm and cyberflashing, and the disabling comments measure to serious self-harm.[326] This is reflected in measures ICU J1 and ICU J2.

How these measures work

4.11 Full details of how the user blocking[327] and muting[328] (ICU J1) and disabling comments (ICU J2) measures work are available in the December 2024 Statement.[329] This includes details on how measure ICU J1 is designed to offer users the option to block individual connected or unconnected user accounts, or all unconnected user accounts and to offer users the option to mute other individual user accounts on the service. It also explains how service providers in scope of measure ICU J2 should offer every registered user the option of disabling comments on their own posts, to prevent other users from commenting on their content.

4.12 For providers brought into scope through this extension, providers may either:

a) offer the controls to users who have not been determined to be adults using highly effective age assurance; or
b) offer the controls to all users on child-accessible parts of the service.

4.13 This approach is intended to ensure that children benefit from the measures, while giving providers flexibility in how they are applied. Providers of large user-to-user services with the relevant risks and functionalities should continue to offer these controls to all users.[330]

---

[325] We have made some amendments to the formatting of the measures, to enhance readability. These amendments do not change the substance of the measures. In particular, we have incorporated the definition of 'registered relevant user' into the defined term 'applicable user'. We have also reframed the definition of 'applicable user' so that it focusses on users who have not been assessed to be adults using highly effective age assurance, rather than users who have assessed to be children. This is more consistent with the approach taken elsewhere in Codes.

[326] Our June 2026 New Priority Offences Statement explains our decision to combine the offence of encouraging or assisting suicide and the offence of encouraging or assisting serious self-harm into a single kind of illegal harm: suicide and serious self-harm. We have included cyberflashing as a separate kind of illegal harm.

[327] Blocking means that the users should not encounter each other on the service.

[328] Muting is a 'softer option' than blocking as it allows the muted user to see the muting user's content and leaves channels of communication open.

[329] Ofcom (2024), December 2024 Statement, Vol.2, chapter 12: User controls.

[330] Where providers choose to use highly effective age assurance, they should follow recommendation ICU B1 in the Illegal Content Codes.

Our reasoning

Effectiveness and benefits

4.14 As explained in our December 2024 Statement, offences such as coercive or controlling behaviour, harassment, stalking, threats and abuse, hate, grooming and encouraging or assisting suicide can cause serious harm, as outlined in relevant Illegal Harms Register of Risks (Illegal Harms Register) chapters. These harms are prevalent online. For example, as of January 2026, Ofcom's Online Experiences Tracker showed that 1 in 4 (26%) of UK adult internet users (18+) had seen or experienced content in the past four weeks that they considered to be hateful, offensive, or discriminatory, and which targeted a group or person based on specific characteristics such as race, religion, sex, sexual orientation, disability or gender identity. In addition, 11% of adults surveyed reported they had witnessed "one-off abusive behaviour or threats" in the past four weeks; 3% of UK adult internet users reported seeing or experiencing content "encouraging or assisting suicide" in the past four weeks.[331] In addition, the scale of online grooming is extensive, with the National Society for the Prevention of Cruelty to Children (NSPCC) reporting over 34,000 cases of online grooming crimes against children between 2017 and 2023, and 7,263 cases in 2024 alone.[332]

4.15 These offences can have significant effects on their victims and survivors. Each offence can have negative psychological impacts, including loss of confidence, anger and aggression, increased feelings of self-blame and shame, and lack of personal trust and isolation. Additionally, some offences, such as hate and coercive and controlling behaviour, can lead to financial harm, while others, such as encouraging or assisting suicide, can lead to physical harm and death. More details about the effects of each harm on their victims and survivors can be found in the Illegal Harms Register.[333]

4.16 These harms take place on a variety of service types, and we know that a wide range of functionalities – including direct messaging, posting content, and commenting on content – can be risk factors.[334]

4.17 In our December 2024 Statement (paragraphs 12.61 to 12.82 in volume 2: Service Design and User Choice, chapter 12: User controls), we explain the effectiveness and benefits of these measures in tackling the harms.[335] This includes evidence on how user controls such as blocking, muting, and disabling comments functionalities are widely used by users to protect themselves from harm.

4.18 In our April 2025 Consultation section titled 'Impact assessment', we assessed the case for expanding the scope of these measures.

4.19 Given the prevalence of these harms, the severity of the impact they have and the fact that our analysis suggests that the measures under consideration would be effective in combatting them, we considered that expanding the application of these measures as

---

[331] Ofcom (2026), Online Experiences Tracker – Wave 9 (Fieldwork took place in January 2026).
[332] NSPCC, 2025, "Data shows how criminals are using private messaging platforms to manipulate and groom children".
[333] Ofcom (2024), Illegal Harms Register of Risks (Illegal Harms Register), chapter 5: Controlling or coercive behaviours (CCB); chapter 4: Harassment, stalking, threats and abuse; chapter 3: Hate; chapter 2a: Grooming; chapter 15: Encouraging or assisting suicide (or attempted suicide).
[334] Ofcom (2024), Risk Assessment Guidance and Risk Profiles.
[335] Ofcom (2024), December 2024 Statement, Vol.2 chapter 12: User controls.

proposed would deliver significant incremental benefits to children. The measures would significantly reduce children's exposure to illegal harms, empowering them to determine the content they see on services, who can contact them and who can interact with them. They would also be provided with clear information to help them understand what would happen if they took such actions.

4.20 This assessment is reinforced by evidence that stakeholders provided in the April 2025 Consultation.[336]

4.21 C3P highlights Thorn's research showing that blocking is the most common response among nine to 17-year-olds to potentially harmful online experiences, followed by reporting and muting. This underscores the critical role of blocking and muting in helping children manage online risks.[337]

4.22 Earlier this year, the UK Government announced plans to introduce new legislation to restrict under-16s' access to social media, restrict their access to certain harmful functions on a wider range of services, and restrict those functionalities by default for 16- and 17-year-olds.

4.23 Notwithstanding these restrictions, our assessment is that child users will still benefit from our user controls measures being extended as proposed. Following the implementation of the Government's proposed approach, 16- and 17-year-olds will maintain access to services with the relevant functionalities and they will experience the risk of harm discussed above, although certain functionalities will be restricted by default. Children will also continue to have access to some services and functionalities outside the scope of the relevant restrictions. In our view, these measures remain important to ensure protection of all children online.

4.24 We therefore remain of the view that expanding the scope of the blocking and muting measures will deliver significant benefits. A number of stakeholder responses also supported our assessment that the provision of accessible information about the tools will deliver significant benefits. The Scottish Government and a range of civil society organisations provided feedback on the importance of accessible and inclusive information for these measures.[338] The Scottish Government also stated that providers must clearly inform children and young people of their online rights and in an age-appropriate way.[339]

4.25 We also received feedback from some stakeholders who disagreed with our measures, but none of them presented evidence that warranted changes to the extension or called into question our assessments of its benefits or effectiveness. We address this feedback in the relevant chapter in annex 1: Further stakeholder responses.

---

[336] Stakeholders who provided evidence include: Antisemitism Policy Trust response to the April 2025 Consultation, pp.1-3; Canadian Centre for Child Protection (C3P) response to the April 2025 Consultation, pp.2-6.; Common Sense Media response to the April 2025 User Controls Consultation, pp.2-5; Internet Matters response to the April 2025 Consultation, pp.1-2.

[337] C3P response to the April 2025 Consultation, p.3.

[338] Antisemitism Policy Trust response to the April 2025 Consultation, p.5.; Parent Zone response to the April 2025 Consultation, pp.3-4.; Samaritans response to the April 2025 Consultation, p.1.; Scottish Government response to the April 2025 Consultation, p.2.

[339] Scottish Government response to the April 2025 Consultation, pp.1-2.

4.26 While stakeholder responses have not led us to fundamentally revise our assessment of the benefits of these amendments, several stakeholders provided feedback on expanding the scope to include adults and children. We discuss this in the following sub-section.

Expanding the scope to include adults and children

4.27 In our December 2024 Statement, we outlined that while we consider these measures would benefit all users, they would be especially helpful for children by playing an important role in helping users to avoid harms such as harassment, stalking, threats and abuse, grooming, hate, encouraging or assisting suicide and coercive and controlling behaviour. These reasons also supported extending the proposed measures to better protect children from the specific illegal harms identified.

4.28 In response to our April 2025 Consultation, some stakeholders argued the proposed scope expansion should be rolled out to both adults and children. This included Samaritans and VoiceBox.[340] Their main argument was around the importance of extending protection to all users (notably from suicide and self-harm content). They stated this would likely simplify compliance for providers, as it would be cheaper for service providers to rollout, and would avoid the perceived complexity and privacy issues around highly effective age assurance. VoiceBox added specifically that it would avoid driving children to riskier services or using virtual private networks (VPNs) to avoid highly effective age assurance.[341]

4.29 The Scottish Government said these measures could also help vulnerable adults. It supported our plan to collect evidence on how well the measures protect users from illegal harm, to guide any future expansion.[342]

4.30 In the light of these responses, we considered again whether it would be proportionate to include adults in the expanded application of the measures. We reviewed the evidence of harm available to us and considered children's higher levels of vulnerability and the need for a higher level of protection for children than adults.[343] We also noted that our user-to-user Codes of Practice must be compatible with the objective of providing a higher standard of protection for children than for adults.[344]

4.31 In judging how far to extend the measures, we must make a finely balanced regulatory judgement which weighs, on the one hand, the benefits the expansion would deliver to users and, on the other hand, the burden on small service providers. Given the concerns service providers have raised about impact and costs, we are not satisfied that it would be proportionate to expand the measures in such a way as to capture adult users on services within scope of the expansion. However, given the evidence of children's greater vulnerability to these harms, the objective in the Act that services are designed and operated so as to give children a higher standard of protection and the fact that service providers in scope of the expansion will already need to apply a similar measure to follow the Protection of Children Code of Practice for user-to-user services (Protection of Children Code), we are satisfied that focusing the expansion on child accounts is proportionate. We

[340] Samaritans response to the April 2025 Consultation, p.1; VoiceBox response to the April 2025 Consultation, p.2.
[341] VoiceBox response to the April 2025 Consultation, pp.2-3.
[342] Scottish Government response to the April 2025 Consultation, p.2.
[343] Section 3(4A)(b) of the Communications Act 2003.
[344] Paragraph 4(a)(vi) of Schedule 4 to the Online Safety Act 2023 (the Act).

also continue to monitor emerging evidence to assess whether there may be a case for bringing adult users in scope of the expanded application of the measures.

4.32 Although the extension is primarily intended to ensure children receive these protections, providers may apply the measures to all users. Indeed, where providers do not use highly effective age assurance to determine which users are adults, the measures should be offered to all users of the service.

4.33 Regarding the second point that it would be more cost efficient to apply the measures to all users, providers may choose to apply the measures to all users if they indeed deem it to be more cost effective. See the section 'Impact and costs' for further details.

4.34 We have therefore not amended the design of the measures in the light of the stakeholder responses set out above.

Impacts and costs

4.35 This section assesses the direct and indirect costs associated with service providers implementing and maintaining these user control tools where they are not currently offered.

4.36 Our assessment of the impact on service providers is consistent with both the December 2024 Statement (volume 2, Section 12) and the April 2025 Statement (volume 3, Section 18) for these measures, which can be referred to for more information.

4.37 In our April 2025 Consultation, we noted that the number of service providers impacted by the extension of these measures is likely to be quite small. Only those that are not already in scope of the measures recommending these controls in the current Illegal Content Codes (measures ICU J1 and ICU J2) or the Protection of Children Code (measures PCU J1 and PCU J2) would be brought into scope.[345] We expect only a small number of service providers to be brought into scope because the Protection of Children Code already recommends these measures for providers of smaller services likely to be accessed by children with risks of certain types of content harmful to children, and there are similarities between those harms and the illegal harms that would bring services into scope of these measures.

Direct costs

4.38 In our April 2025 Consultation we estimated the costs that a relevant service provider will incur for the implementation of the user control measures.[346]

4.39 In Table 4.1, we present our updated cost estimates which use the latest salary data from the Office for National Statistics (ONS).[347] The cost ranges reflect the considerable variation across service providers due to different systems and functionalities. In some cases, service providers may already have the measures (or parts of them) in place, meaning costs may be lower.

---

[345] These are expected to be providers of smaller services that are likely to be accessed by children that identify a risk of relevant types of illegal content on their service, but have not identified a relevant risk of the types of content specified in measures PCU J1 and/or PCU J2 in the Protection of Children Code.
[346] Ofcom (2024), December 2024 Statement, Vol.2, chapter 12: User controls.
[347] We assume implementation of the user blocking and muting measure would require 20 to 150 days of software engineering time, matched with equal time from professional occupational staff. The equivalent assumptions are 5 to 50 days of software engineering time matched with professional occupation time for the disabling comments measure, and 0.5 to 10 days for making these measures known to users. See annex 5 for additional details on our economic assumptions and analysis.

Table 4.1: Estimated costs to service providers of the user controls measures

Measure	One-off implementation cost	Annual maintenance cost
User blocking and muting	£11,000 to £170,000	£3,000 to £43,000
Disabling comments	£3,000 to £57,000	£700 to £14,000
Making measures known to users	£300 to £11,000	£70 to £3,000

Direct costs related to age assurance

4.40 We expect that some providers will use highly effective age assurance to determine which users are not adults and apply the measures to those users, while others will apply the measures to all users. If a provider uses highly effective age assurance to target the measures at such users, they may incur costs relating to its use. They may also incur costs relating to its implementation if their service does not already have highly effectively age assurance and they choose to implement it to offer these control features only to users who have not determined to be adults. Alternatively, service providers can apply the measures to all users and not incur any costs related to highly effective age assurance.

4.41 Services can choose between developing highly effective age assurance internally or using a third-party provider. We have estimated the costs of using an external provider, which are typically linked to the volume of age checks conducted and can range from £0.05 to £0.30 per age check. We set out indicative age assurance costs for providers of services of different sizes in annex 5: Further detail on economic assumptions and analysis. For example, for a service with 100,000 users and 10,000 new users each year, there could be a cost of £5,000 to £30,000 to conduct age assurance checks for all existing users, and a cost of £1,000 to £3,000 annually for new users. For a service with 7,000,000 users and 70,000 new users each year, the one-off costs could be £350,000 to £2,100,000, and annual costs of £4,000 to £21,000. We note that providers that already have highly effective age assurance, or have been recommended to implement it as part of other measures, will have already incurred most of this cost.

4.42 We did not receive stakeholder feedback relating to our assessment of direct cost for service providers and so our approach to costs remains unchanged from our position in the consultation.

Indirect costs

4.43 In our April 2025 Consultation we recognised the potential for these measures to have an indirect impact on service providers. They could reduce overall engagement on services where user blocking or disabling comments substantially alters users' experiences of a site or community, which would vary based on how the service works. We also identified potential countervailing positive effects; the measures may prevent some users from leaving or disengaging with a service by making users feel safer online.[348]

4.44 We did not receive stakeholder feedback relating to our assessment of indirect costs for service providers and so our assessment remains unchanged from our position in the consultation.

[348] Ofcom (2024), December 2024 Statement, Vol.2, chapter 12: User controls.

Rights impact

4.45 We have considered the rights impact for the extension of these measures together given the assessment is similar. We have noted where we consider there to be differences.

4.46 Providers of services which fall in scope of the extended measures may use highly effective age assurance to apply them only to children, otherwise they should apply them to all users. We have therefore considered the impact on both child and adult users.

Freedom of expression and freedom of association

4.47 The rights to free expression and association include the right to receive and impart information, and to associate with others. Any restriction of these rights must pursue a legitimate aim, be prescribed by law and be proportionate.

4.48 We did not receive any stakeholder feedback on our assessment of the impact of our proposals on these rights.

4.49 We recognise that these measures will interfere with users' ability to receive or impart information, and their freedom to associate, where they are blocked or muted or are unable to comment on content. However, we consider any such interference will be limited as these users will still be able to post content on other parts of the service and experience no limitation on their ability to associate with users who have not chosen to restrict their interaction with them.

4.50 Users who choose to block or mute other users and/or disable comments will be exercising their rights to freedom of expression and association by choosing with whom they impart information and associate. These measures will therefore have a positive effect on these rights by empowering users to interact with other users only if they choose to.

4.51 By effectively enabling users to protect themselves from illegal content, any interference arising from the expansion of these measures pursues the legitimate aims of preventing crime, protecting health and morals, and protecting the rights of others. Any interference with user rights will be in accordance with the law, as these measures support providers in complying with their duties under section 10 of the Act to protect users from illegal content, and the extended application of the measures specifically seeks to offer children a higher level of protection as required by the Act. They are prescribed by law, as the measures are clear in recommending that providers offer users the option to block or mute other users, and to disable comments on their content, setting out the effect this should have on different functionalities and features. Given the limited impact on user rights and the clear evidence that these measures are effective in protecting users from harm, we are satisfied that any interference is proportionate.

4.52 For the same reasons, we are satisfied that any impact on service providers' freedom of expression will also be limited, justified and proportionate.

Privacy

4.53 Any interference with users' right to privacy must also pursue a legitimate aim, be prescribed by law and be proportionate.

4.54 techUK, which also separately stated that it supports the objectives of the measure, emphasised the importance of privacy-respecting methods of age assurance that protect user data.[349] Two stakeholders raised concerns about the impact on user privacy of highly

[349] techUK response to the April 2025 Consultation, p.2-3; techUK email dated 12 October 2026.

effective age assurance.[350] We acknowledge that if providers use highly effective age assurance to target these measures at users who have not been determined to be adults, there could be privacy impacts associated with its use. However, we expect that providers will only apply highly effective age assurance for the purposes of these measures where they are already deploying it for other reasons. They should also follow Recommendation ICU B1 in the Illegal Content Codes, which acts as a safeguard for user privacy. Any additional privacy or data protection impacts as a result of the use of highly effective age assurance in connection with these measures will be very limited.

4.55 We consider that any interference with users' right to privacy arising from these measures is likely to be extremely limited, for example providers processing user data to give effect to a block or a mute. On the other hand, blocking and muting may have benefits for users' privacy by giving them additional options on who to share their personal information with.

4.56 To the extent there is any interference with user privacy, this will pursue the legitimate aims and be prescribed by law for the reasons set out in paragraph 4.47. Noting the effectiveness of these measures in protecting users from harm, we are satisfied that any interference will be proportionate.

4.57 Where providers are required to process users' personal data to implement these measures, they must comply with their obligations under data protection law, including having regard to the data protection principles of fairness, transparency, accuracy and data minimisation. Providers should refer to relevant ICO guidance where appropriate.[351]

Who these measures apply to

4.58 In our April 2025 Consultation we proposed that, in addition to those services already in scope of these measures in the Illegal Content Codes, it would be proportionate to recommend the measures apply to providers of smaller services which are likely to be accessed by children that are either high risk for the relevant illegal harms, or medium risk for the relevant illegal harms and have over 700,000 monthly UK users.

4.59 Responses to the April 2025 Consultation showed broad support for the expansion of the scope of the measures by a range of stakeholders.[352] In the rest of the section, we will explore feedback received on specific areas, including the size of service threshold, impact of the measures on providers of small services, and risks of small services.

4.60 Some respondents argued that we should set the threshold lower and target the measures at services with fewer than 700,000 users.[353]

4.61 We received feedback expressing concerns about the impact the expansion of the measures would have on providers of small services.[354] However, this was contested by other

[350] [✂]; VoiceBox response to the April 2025 Consultation, p.2.
[351] ICO (n.d.), UK GDPR guidance and resources.
[352] 5Rights Foundation response to the April 2025 Consultation, p.1; Antisemitism Policy Trust response to the April 2025 Consultation, p.4; Common Sense Media response to the April 2025 Consultation, p.2; Federated Hermes response to the April 2025 Consultation, p.2; Internet Matters response to the April 2025 Consultation, p.2; [✂]; Online Dating and Discovery Association (ODDA) response to the April 2025 Consultation, p.1; Samaritans response to the April 2025 Consultation, p.2; Scottish Government response to the April 2025 Consultation, p.2; VoiceBox response to the April 2025 Consultation, p.2.
[353] Antisemitism Policy Trust response to the April 2025 Consultation, p.3; C3P response to the April 2025 Consultation, pp.4-5.
[354] Association for UK Interactive Entertainment (Ukie) response to the April 2025 Consultation, pp.3-4; ODDA response to the April 2025 Consultation, p.1.

stakeholders who challenged the idea that small service providers may struggle to implement the measures or do so without harming user experience.[355]

4.62 We have considered this feedback and have decided that our proposal strikes the right balance between providing increased protection for child users where it is most needed while balancing the impacts of the measures on service providers and unintended impacts on users.

4.63 We recognise the risks on smaller services, and so we are extending the scope of our measures to include providers of all services which are high risk regardless of size, and medium risk services with over 700,000 users. By extending these measures to providers of certain services with no more than seven million users, we are capturing additional services which pose significant risks of harm to children.[356]

4.64 On high-risk services that are not in scope of the equivalent Protection of Children measures, we expect that the absence of these tools could lead to significant unaddressed harm which is potentially severe. We therefore consider the tools proportionate regardless of the size of the service, even if providers of those services are smaller and face comparatively higher costs of implementation.

4.65 For medium risk services, we consider it is proportionate to apply to providers with over 700,000 monthly UK users. This size threshold is expected to capture a wide range of services which are popular among children in the UK which have no more than seven million monthly UK users. We consider that these measures will provide material benefits to users on these services, and that these providers will be able to manage the impacts of the measures on their services.

4.66 For providers of the smallest services with fewer than 700,000 monthly UK users and a medium risk of harm, we have concluded that they would struggle to implement the measures in a way that increases protection from illegal content without material adverse effects for users. In particular, smaller service providers will have more limited resources, and the implementation of these measures may be challenging and, in some cases, require a significant overhaul of the service's systems and architecture. As a result, we are confirming that these measures will not apply to providers of services with fewer than 700,000 monthly UK users that are medium or low risk for illegal harms. Our assessment is that providers of the smallest services that are not high risk would meet their Illegal Content Duties more effectively by focusing on the other measures recommended in our Illegal Content Codes which will help to reduce the risk of harm.

4.67 We have therefore decided that, in addition to those services already in scope of these measures in the Illegal Content Codes, it is proportionate to recommend these measures for the following service providers:

- Measure on blocking and muting (ICU J1): Providers of user-to-user services likely to be accessed by children that have user profiles, have at least one of the following functionalities (user connections, posting content, user communication), and have:

[355] C3P response to the April 2025 Consultation, p.4; Parent Zone response to the April 2025 Consultation, pp.4-5.
[356] We have amended the wording from our consultation proposal, from "below" to "no more than" seven million monthly active UK users, to ensure that services with seven million users are in scope of the measures.

- no more than seven million monthly UK users and are at high risk for at least one of the following kinds of illegal harm: coercive or controlling behaviour; harassment, stalking, threats and abuse; hate; grooming; encouraging or assisting suicide (or attempted suicide) and serious self-harm; cyberflashing; or
- between 700,000 and seven million monthly UK users and are at medium risk for at least one of the relevant kinds of illegal harm.

- Measure on disabling comments (ICU J2): Providers of user-to-user services likely to be accessed by children that have the functionality of commenting on content, and have:
  - no more than seven million monthly UK users and are at high risk for at least one of the following kinds of illegal harm: harassment, stalking, threats and abuse; hate; grooming; encouraging or assisting suicide (or attempted suicide) and serious self-harm; or
  - between 700,000 and seven million monthly UK users and are at medium risk for at least one of the relevant kinds of illegal harm (as above) and have over 700,000 monthly UK users.

Conclusion

4.68 Our analysis shows that the expansion of these measures is likely to deliver significant protections to children from the relevant harms, and that the impacts on services and any consequential impact on rights are proportionate. In practice, we expect this decision to affect very few service providers, limited only to those that are not already in scope of the existing measures in the current Illegal Content Codes or the Protection of Children Code.

4.69 We have decided to proceed with the measures and the extension as proposed in our April 2025 Consultation. The full text of the measures can be found in our Illegal Content Codes of Practice and are referred to as ICU J1 and ICU J2. The measures are part of our Illegal Content Codes of Practice for user-to-user services on CSEA and other duties.

5. Highly effective age assurance in the Illegal Content Codes

What is this chapter about?

Under the Act, age assurance must be highly effective at correctly determining whether a user is an adult or a child. This chapter sets out our decisions to establish a consistent approach to highly effective age assurance across the Illegal Content and Protection of Children Codes, makes minor amendments to the Part 3 HEAA Guidance, and introduces measures on age-assessment appeals.

The measures support protections against serious illegal harms affecting children, including grooming and child sexual exploitation and abuse (CSEA), by helping providers accurately determine which UK users are adults and apply appropriate safeguards to all remaining users.

The measures enable providers to use highly effective age assurance to target child-specific protections more effectively, while requiring safeguards, transparency, and appeals processes to ensure age-assurance systems are accurate, fair and proportionate.

What decisions have we made?

We have decided to proceed with the proposed measures. Our final Code measures are summarised in the table below:

Number in Codes	Recommended measure	Who should implement this
ICU B1	Use an age assurance process that meets the highly effective age assurance criteria (technical accuracy, robustness, reliability and fairness) and have regard to key implementation principles.	Providers of user-to-user services that use an age assurance process to determine which UK users are child users.
ICU D15	Provide an age-assessment appeals process and take appropriate action where an appeal is upheld.	Providers of user-to-user services that use an age assurance process to determine which UK users are child users and that are large or multi-risk services.
ICU D16	Provide an age-assessment appeals process and take appropriate action where an appeal is upheld.	Providers of user-to-user services that use an age assurance process to determine which UK users are child users and that are neither large nor multi-risk services.

Number in Codes	Recommended measure	Who should implement this
PCU B1	Amend the Protection of Children Codes to codify the highly effective age assurance criteria within the measure and align its structure with the Illegal Content Code.	Providers of user-to-user services likely to be accessed by children that use an age assurance process to determine which UK users are child users.

Why have we made these decisions?

These decisions will ensure providers that use age assurance have effective and consistent systems in place to determine which users are adults, enabling them to apply child-specific protections more accurately and consistently to all remaining users. This supports measures to prevent illegal harms affecting children, including grooming and CSEA, by ensuring that child-specific protections are provided for child users.

The measures are also designed to ensure that age assurance processes are accurate and transparent, and provide users with appropriate redress where their age has been incorrectly assessed. This should ensure that age assurance systems are fair and proportionate.

Introduction

5.1 In our Age Assurance and Children's Access Statement (January 2025 Statement),[357] we confirmed our criteria-based, technology-neutral and future-proofed approach to highly effective age assurance, as reflected in our draft Guidance on Highly Effective Age Assurance for Part 3 Services (Part 3 HEAA Guidance).[358] We also stated that we "expect the approach to highly effective age assurance that we are setting out today to apply to all parts of the online safety regime."[359]

5.2 In our June 2025 Additional Safety Measures Consultation (June 2025 Consultation), we set out proposals to use highly effective age assurance to support our Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes). Specifically, we proposed that service providers can use highly effective age assurance to implement measures that protect children from grooming and other illegal harms. We set out the detail of the measures for which we recommend the use of highly effective age assurance in the relevant chapters.

5.3 The measures discussed in this chapter support the implementation of highly effective age assurance across a number of recommended measures elsewhere in this statement, including:

- Chapter 4: Extending user controls (ICU J1 and J2), where providers may use highly effective age assurance when making certain user controls available to users who have not been determined to be adults.

[357] Ofcom (2025), January 2025 Statement.
[358] Ofcom (2025), Guidance on highly effective age assurance (Part 3 HEAA Guidance); Ofcom (2025), Guidance on highly effective age assurance and other Part 5 duties.
[359] Ofcom (2025), January 2025 Statement, p.3.

- Chapter 6: Increasing the effectiveness of safety defaults and supportive messaging (ICU F1 and F2), where providers may use highly effective age assurance to apply protections against grooming to users who have not been determined to be adults.
- Chapter 11: Recommender systems (ICU E2 and ICU E3), where providers may use highly effective age assurance to apply recommender system protections to users who have not been determined to be adults.

5.4 This chapter sets out the overarching framework for highly effective age assurance, associated appeals measures, and amendments to the Part 3 HEAA Guidance that support the implementation of these measures across the Illegal Content Codes. Service providers should also refer to the Part 3 HEAA Guidance for further detail on what constitutes highly effective age assurance and how it should be implemented.[360]

5.5 In this chapter, we address the June 2025 Consultation feedback on our overall approach to highly effective age assurance in the Illegal Content Codes and set out our final decision in respect of:

a) establishing a consistent approach to highly effective age assurance in the Illegal Content Codes and minor amendments to the Part 3 HEAA Guidance; and
b) measures on appropriate action for age assessment appeals.

5.6 Earlier this year, the Government announced that the use of highly effective age assurance will be critical to the successful implementation of age restrictions for social media services. We recognise that this would introduce a new age limit of 16 alongside 18. We recently published a rapid assessment of highly effective age assurance for determining whether someone is over 16 to help inform parliamentary debate on these age restrictions.[361] However, for the purposes of recommended measures on the implementation of highly effective age assurance in certain circumstances, our current approach to age assurance continues to apply.

Establishing a consistent approach to highly effective age assurance in the Illegal Content user-to-user Codes and minor amendments to the Part 3 HEAA Guidance

5.7 In the June 2025 Consultation, we proposed:

a) Introducing a measure that (1) defines highly effective age assurance for the purposes of the Illegal Content Codes, so that it is consistent with the definition used in the Protection of Children Code of Practice for user-to-user services (Protection of Children Code); and (2) sets out principles that service providers should have regard to when using an age assurance process.
b) Changing the way in which we present the definition of highly effective age assurance in the Protection of Children Code and to mirror this presentation in the Illegal Content Codes.

[360] Ofcom (2025), Guidance on highly effective age assurance.
[361] Ofcom (2026), Rapid assessment of highly effective age assurance at 16.

c) Amending Part 3 HEAA Guidance to broaden its scope to cover highly effective age assurance used in connection with the Illegal Content Codes.

Our decision

5.8 Having considered the consultation feedback, we have decided to confirm the proposed amendments to the Illegal Content Codes, Protection of Children Code and the Part 3 HEAA Guidance. We set out our reasoning for confirming our approach in the following sections.

5.9 The full text of the measures can be found in our Illegal Content and Protection of Children user-to-user Codes of Practice and are referred to as ICU B1 and PCU B1.

5.10 The Part 3 HEAA Guidance now also includes our proposed amendments. Service providers should refer to the Part 3 HEAA Guidance for further detail on what constitutes highly effective age assurance and how it should be implemented.[362]

How this approach works

5.11 As explained in our June 2025 Consultation, there are three elements to establishing a consistent approach to highly effective age assurance in the Illegal Content Codes and the Part 3 HEAA Guidance:

i) defining highly effective age assurance;
ii) introducing a measure which sets out principles that providers should have regard to when implementing an age assurance process; and
iii) changing the way in which the definition is presented in the Protection of Children Code and reflecting this presentation in the Illegal Content Codes.

5.12 We set out how our approach works for each of these proposals below.

Defining highly effective age assurance

5.13 We define highly effective age assurance in the Illegal Content Codes, to provide clarity on how to ensure age assurance is highly effective for the purpose of the Illegal Content Codes and so that it is consistent with the definition used in the Protection of Children Code and our Part 3 HEAA Guidance. The definition sets out that highly effective age assurance is an age assurance process that fulfils the criteria of technical accuracy, robustness, reliability, and fairness.

Principles for implementing age assurance

5.14 We have also introduced a measure into the Illegal Content Codes which sets out principles that service providers should have regard to when implementing highly effective age assurance where it is recommended to support other measures. This reflects the recommendation in measure PCU B1 in the Protection of Children Code. The measure will provide safeguards to protect users' rights to freedom of expression and privacy (including data protection).

5.15 The principles are:

a) that age assurance should be easy to use, including by children in the UK of different ages and with different needs;

[362] Ofcom (2025), Guidance on highly effective age assurance.

b) that age assurance should work effectively for all users, regardless of their characteristics or whether they are members of a certain group;
c) that providers should consider how different kinds of age assurance will work together;
d) that providers should consider the age-appropriate design code and the Information Commission's Office (ICO) Opinion entitled 'Age Assurance for the Children's Code' published on 18 January 2024.

5.16 The measure also recommends that providers should ensure users can easily access information explaining the purpose and operation of the age-assurance process before commencing it.

5.17 Lastly, the measure recommends that providers should not publish content that directs or encourages users to circumvent the age assurance process or access controls used on the service.

Changing the presentation of the definition in the Protection of Children Code and reflecting this in the Illegal Content Codes

5.18 We have changed how the definition of highly effective age assurance is presented in the Protection of Children Code, without altering its substance.

5.19 The definition of highly effective age assurance was set out in the "Definitions and interpretation" section of the Protection of Children Code and is not itself codified within the recommended age assurance measures.[363] [364]

5.20 For the Illegal Content Codes, we took a different approach by codifying this definition within a specific measure. The relevant measure details the criteria that a service provider's age assurance process should meet for it to be considered highly effective.

5.21 We amended the Protection of Children Code to adopt this approach. We moved the highly effective age assurance definition from section 5 ('definitions and interpretations') to measure PCU B1, which set out the principles which providers must have regard to when implementing an age assurance process. This approach ensures consistency and provides clarity to service providers.

5.22 We also replaced the term "identify" with "determine". This more accurately reflects our policy intent and ensures consistency with terminology used elsewhere in the Protection of Children Code (for example, references to a user being "determined to be an adult through the use of highly effective age assurance").

Minor amendments to the Part 3 HEAA Guidance

5.23 Service providers in scope of Part 3 of the Online Safety Act 2023 (the Act) should refer to the Part 3 HEAA Guidance to understand what constitutes highly effective age assurance and how it should be implemented.[365]

5.24 We are now recommending highly effective age assurance for the purpose of carrying out certain measures in the Illegal Content Codes. Therefore, we amended the Part 3 HEAA Guidance to reflect an expansion of scope by adding references to the Illegal Content Codes and to reflect any changes to the wording and numbering of the relevant measures. These amendments ensure the Part 3 HEAA Guidance can be relied on by Part 3 services when

[363] Protection of Children Code of Practice for user-to-user services (Protection of Children Code), p.87.
[364] Protection of Children Code, Section 4, PCU B1
[365] Ofcom (2025), Guidance on highly effective age assurance.

implementing relevant measures in both the Protection of Children and Illegal Content Codes as they evolve.

5.25 The updated Part 3 HEAA Guidance was published alongside this statement.

Our reasoning

Effectiveness and benefits

5.26 We consider these amendments will support the Illegal Content Codes by establishing a clear and effective framework for the use of highly effective age assurance. Aligning the approach to age assurance across the Illegal Content and Protection of Children Codes provides regulatory coherence and consistency. It also reduces implementation complexity and costs for service providers by enabling them to implement a single, consistent age assurance process.

5.27 In particular, the amendments set out clear principles for service providers implementing highly effective age assurance which will help ensure it is easy to use, effective for all users regardless of their characteristics, and has regard to interoperability. It also promotes an approach that protects users' rights to freedom of expression and privacy (including data protection), while ensuring that users not determined to be adults receive appropriate protections, alongside supporting transparency and fairness.

5.28 We recognise that highly effective age assurance was set out and defined within the Protection of Children Code, but the definition was not set out with the recommendations to implement an age assurance process. We consider that moving the definition into the operative code measures makes its purpose and requirements clearer for service providers and, as a result, promote compliance.

Evidence for the effectiveness of highly effective age assurance

5.29 In response to the June 2025 Consultation, the British and Irish Law Education Technology Association (BILETA) recommended waiting to assess highly effective age assurance's real-world impact before expanding it to other codes to avoid undermining the Act's long-term aims.[366] It argued that it is too early to judge the effectiveness of highly effective age assurance, noting the Protection of Children measures (which recommend highly effective age assurance for compliance with the Children Safety Duties) have only been in force since July 2025 and evidence remains limited. It pointed to Ipsos public opinion research from 2025, which suggests scepticism about the effectiveness of age checks across the UK population, as indication that the evidence base is still developing.[367] Two individual respondents also questioned whether the current evidence supported the effectiveness of highly effective age assurance.[368]

5.30 We addressed feedback on the effectiveness of highly effective age assurance in our January 2025 Statement and our April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement). There are a range of age assurance methods which are capable of being highly effective, giving us confidence that our overall criteria-based approach to age assurance is fit for purpose. Further findings on the use and effectiveness

[366] British and Irish Law Education Technology Association (BILETA) response to the June 2025 Consultation, p.31.
[367] Ipsos, 2025. 'Britons Back Online Safety Act's Age Checks, but Are Sceptical of Effectiveness and Unwilling to Share ID'.
[368] [✂]; We are Survivors response to the June 2025 Consultation, p.8.

of age assurance methods during the early months of the child safety duties being in force are set out in our age assurance report.[369] We have also gathered evidence directly from children and parents on their experiences of age assurance and the Protection of Children measures following implementation, providing evidence on how these measures are operating in practice.[370]

5.31 We recognise that the evidence base on the effectiveness of highly effective age assurance will continue to develop, and that further evidence may reinforce or challenge our current assessment over time. Nevertheless, based on the evidence currently available, we continue to consider our approach to age assurance to be effective in enabling child-specific protections. We have also weighed the risk of delaying protections for children against the risk that highly effective age assurance may not operate in all cases as intended and consider that the former outweighs the latter.

5.32 This demonstrates that our approach will effectively deliver meaningful benefits for children by providing a clear, coherent, and proportionate framework for highly effective age assurance.

The use of highly effective age assurance in the Illegal Content Codes

5.33 Stakeholders also raised a range of views on the proposed use of highly effective age assurance in the Illegal Content Codes.

5.34 The Computer & Communications Industry Association (CCIA) and Meta Platforms Inc. (Meta) questioned recommending the use of highly effective age assurance to comply with illegal content safety duties that apply to all regulated user-to-user services, rather than specifically to regulated user-to-user services likely to be accessed by children.[371] They, along with X, stressed that highly effective age assurance is required only for preventing children from accessing primary priority content and warned that extending it to areas such as livestreaming would impose unnecessary costs, infringe users' rights, and go beyond the Act's intended purpose.[372]

5.35 Conversely, a range of other stakeholders supported the use of highly effective age assurance to apply child safety measures in the Illegal Content Codes.[373] They emphasised

---

[369] Ofcom (2026), Use of Age Assurance Report.
[370] Ofcom (2026), Experiences of Age Assurance, Use of VPNs, and attitudes to the Protection of Children Measures.
[371] Computer & Communications Industry Association (CCIA) response to the June 2025 Consultation, p.14; Meta Platforms Inc. (Meta) response to the June 2025 Consultation, p.7.
[372] Computer & Communications Industry Association (CCIA) response to the June 2025 Consultation, p.14; Meta response to the June 2025 Consultation, p.7; X response to the June 2025 Consultation, p.2.
[373] A member of the Philippine Survivor Network response to the June 2025 Consultation, pp.21-22; Antisemitism Policy Trust response to the June 2025 Consultation, p.3; Barnardo's response to the June 2025 Consultation, p.16; Canadian Centre for Child Protection (C3P) response to the June 2025 Consultation, p.17; Children's Commissioner's office (CCO) response to the June 2025 Consultation, p.21; Commissioner Designate for Victims of Crime for Northern Ireland response to the June 2025 Consultation, p 4; Department of Justice for Northern Ireland response to the June 2025 Consultation, p.14; Digital Resilience in Education – Welsh Government response to the June 2025 Consultation, p.14; Evangelical Alliance response to the June 2025 Consultation, p.9; ICO response to the June 2025 Consultation, p.29; Internet Matters response to the June 2025 Consultation, p.18; Internet Watch Foundation (IWF) response to the June 2025 Consultation, pp.50-52; Marie Collins Foundation response to the June 2025 Consultation, p.11; [✂]; National Center for Missing & Exploited Children (NCMEC) response to the June 2025 Consultation, pp.7-8; [✂]; National Police Chiefs' Council (NPCC) response to the June 2025 Consultation, p.14; The Online Dating and Discovery Association

that weaker approaches such as self-declaration are ineffective, and that robust age checks are essential for applying child-specific protections, reducing grooming, sexual extortion and other online harms, and ensuring safety measures are targeted at the right users and operate effectively in practice. Stakeholders also welcomed the alignment of highly effective age assurance requirements across the Illegal Content and Protection of Children Codes, arguing that it would improve regulatory clarity, consistency and ease of implementation.

5.36 We acknowledge the additional feedback on the scope and impact of recommending highly effective age assurance for the Illegal Content Codes. In particular, we note that the Act does not exclude the recommendation to use highly effective age assurance in a Code of Practice to meet the overall aims of the illegal content safety duties. Schedule 4 of the Act imposes requirements on Ofcom to have regard for the general principles when designing Codes of Practice[374] and to ensure that measures described in Codes of Practice are compatible with the pursuit of the online safety objectives.[375]

5.37 The use of highly effective age assurance is compatible with the pursuit of the online safety objectives, because it supports services being designed and operated in a way that a higher standard of protection is provided for children than for adults.[376]

5.38 We had regard to the general principles by analysing and determining that the highly effective age assurance measures were proportionate to the risk of harm that we intend for the measure to address.[377] We determined that the measures assist service providers to more effectively mitigate illegal content, which can severely impact children, including CSEA, terrorism, hate and suicide content.

5.39 Though the purpose of the highly effective age assurance measures is to demonstrate compliance with the illegal content safety duties, the focus is on protecting children from severe risks of illegal harm. Therefore, we have taken further steps to have regard to the additional principles that Ofcom is required to consider when deciding to recommend the use of age assurance measures to comply with the children safety duties.[378] This includes balancing the risk levels, nature and severity of harm against the protection of users' rights to freedom of expression, as well as having regard to the effectiveness of age assurance in determining the age of users and to interoperability.

Impact and costs

5.40 We consider the impact and costs associated with confirming a codified definition of highly effective age assurance and accompanying principles in the Illegal Content Codes.

5.41 We expect service providers will incur costs to understand our Part 3 HEAA Guidance and assess their age assurance process against the criteria, as well as having regard to the principles of accessibility and interoperability as part of their implementation.

5.42 Further, as age assurance technology evolves, the service providers may have to review and update their approach to ensure compliance over time.

(ODDA) response to the June 2025 Consultation, p.8; Verifymy response to the June 2025 Consultation, pp.3, 9.

[374] Paragraphs 1 of Schedule 4 to the Online Safety Act 2023 (the Act).

[375] Paragraphs 3 of Schedule 4 to the Act.

[376] Paragraph 4 of Schedule 4 to the Act.

[377] Paragraph 2 of Schedule 4 to the Act.

[378] Paragraph 12 of Schedule 4 to the Act.

5.43 The scale of these costs is likely to be highly dependent on each service provider's context and approach to implementation. But generally, we expect these costs to be limited and, in some cases, overlap with actions service providers would already take to comply with existing obligations and guidance.

5.44 During the consultation, several stakeholders raised concerns that expanding highly effective age assurance could increase costs or complexity for service providers, particularly smaller services with more limited resources.[379] Stakeholders cautioned that disproportionate burdens could affect innovation, competition, and market diversity, and emphasised the importance of proportionate and scalable regulatory expectations.[380]

5.45 The principles accompanying the definition allow providers flexibility to adopt approaches that reflect their service design, user base, and risk profile, helping to ensure that implementation costs remain proportionate.

5.46 Overall, we consider that the costs associated with confirming the definition and principles are likely to be limited and proportionate. These costs are outweighed by the benefits of improved regulatory coherence and consistency across the Illegal Content and Protection of Children Codes.

Rights impact

5.47 This section considers how the amendments and measure affects users' rights, including freedom of expression, privacy, and data protection.

Freedom of expression and freedom of information

5.48 Article 10 of the European Convention on Human Rights (ECHR) sets out the right to freedom of expression, which encompasses the right to hold opinions and to receive and impart information and ideas without unnecessary interference by a public authority. Article 11 sets out the right to associate with others. We must exercise our duties under the Act in light of users' and services' Article 10 and 11 rights and not interfere with these rights unless we are satisfied that to do so is prescribed by law, pursues a legitimate aim, is proportionate to the legitimate aim, and corresponds to a pressing social need.

5.49 Our decisions around highly effective age assurance in both the Protection of Children and Illegal Content Codes are beneficial to users' rights to freedom of expression and association. In particular, the recommendation that service providers have regard to certain principles when implementing highly effective age assurance creates a better experience for users and reduces unnecessary barriers when accessing services, features, and content, while respecting users' rights.

5.50 We recognise that some stakeholders raised concerns about the potential impact of highly effective age assurance on freedom of expression and association.[381] However, these concerns did not include evidence that would call into question our assessment of the impact on users' rights. We have addressed those concerns elsewhere in this statement, including in the relevant chapter in annex 1: Further stakeholder responses.

---

[379] [✂]; ODDA response to the June 2025 Consultation, p.9; [✂]; [✂].
[380] ODDA response to the June 2025 Consultation, p.8; [✂].
[381] ACT – The App Association response to the June 2025 Consultation, p.9; [✂]; Francis, S. response to the June 2025 Consultation, p.12-13; [✂]; Harris, B. response to the June 2025 Consultation, p.7; Name Withheld 4 response to the June 2025 Consultation, p.9; Name Withheld 9 response to the June 2025 Consultation, p.9; NetChoice response to the June 2025 Consultation, p.4; [✂].

Privacy and data protection

5.51 Article 8 of the ECHR sets out the right to respect an individual's private and family life. We must exercise our duties under the Act in light of users' privacy rights and not interfere with these rights unless we are satisfied that to do so is prescribed by law, pursues a legitimate aim, is proportionate to the legitimate aim, and corresponds to a pressing social need.

5.52 The measures ensure a consistent approach to highly effective age assurance by establishing the same definition for highly effective age assurance across all Codes of Practice. The measures also include recommended safeguards, which are designed to help to protect users' rights to privacy, including data protection.

5.53 The Online Safety Network, several individual respondents and civil society stakeholders raised concerns about privacy and data protection, arguing that the proposals did not place sufficient emphasis on these issues and highlighting data security risks associated with highly effective age assurance, including the sensitivity of personal information and the risk of data breaches.[382] Several stakeholders also suggested privacy-preserving alternatives, including interoperable age tokens.[383]

5.54 5Rights Foundation argued that we should make privacy and data protection a clear requirement. It warned that recent data breaches risk undermining trust and urged us to adopt stronger technical privacy standards such as IEEE 2089.12024.[384]

5.55 The Internet Watch Foundation (IWF) added that there is uncertainty about how providers should navigate both Ofcom's highly effective age assurance guidance and the ICO Age-Appropriate Design Code. They, along with techUK, highlighted the need for clearer coordination between regulators to give services confidence in meeting their legal obligations.[385]

5.56 We acknowledge concerns for clarity and emphasis around privacy and data protection. Our decisions are consistent with our Part 3 HEAA Guidance, which set out clearly the expectations on service providers who implement highly effective age assurance to have regard to users' right to privacy, and to consult ICO guidance to understand how to comply with data protection legislation. We expect providers to rely on the Part 3 HEAA Guidance when implementing highly effective age assurance for the Illegal Content Codes.

5.57 We are clear that service providers must comply with data protection legislation when processing users' personal data as part of a highly effective age assurance process and should adhere to the applicable guidance published by the ICO when processing this data, including the Age-Appropriate Design Code.[386] Our decision to establish these existing expectations in our Illegal Content Codes reinforces our position that service providers

---

[382] BILETA response to the June 2025 Consultation, p.33; CCIA response to the June 2025 Consultation, p.14; Edwards, L. response to the June 2025 Consultation, pp.8, 10; [✂]; [✂]; Klaushofer, A. response to the June 2025 Consultation, p.9; Mack, S. (Dr) response to the June 2025 Consultation, p,7-8; Name Withheld 9 response to the June 2025 Consultation, p.9; Online Safety Network response to the June 2025 Consultation, p.25; Open Rights Group response to the June 2025 Consultation, pp.12-13, 39; [✂]; UK Finance response to the June 2025 Consultation, pp.14-15.

[383] Name Withheld 9 response to the June 2025 Consultation, p.9; Name Withheld 11 response to the June 2025 Consultation, p.1; [✂].

[384] 5Rights Foundation response to the June 2025 Consultation, p.13.

[385] IWF response to the June 2025 Consultation, pp. 51, 56; techUK response to the June 2025 Additional Safety Measures Consultation, p.26.

[386] ICO, 2022. Age appropriate design: a code of practice for online services.

implementing age assurance should consider applicable guidance published by the ICO. This position is also reinforced in the Ofcom and the ICO's Joint Statement on Age Assurance,[387] which provides further clarity on how providers can meet their obligations under the Act and UK data protection legislation.

5.58 In the Part 3 HEAA Guidance, we state the importance of providers having regard to privacy rights and data protection requirements, while also confirming that the ICO are responsible for enforcement and oversight of data protection legislation. Our decisions increase the certainty of expectations to take into account and consult relevant guidance from the ICO to assist in complying with data protection law and protecting user's privacy.[388]

Who this measure applies to

5.59 As explained in our June 2025 Consultation, measure ICU B1 in the Illegal Content Codes will apply to user-to-user services that use highly effective age assurance to determine which UK users of the service are child users in order to target illegal harms measures at children.

5.60 The amendments to measure PCU B1 will not affect who the measure applies to.

Conclusion

5.61 We have decided to proceed with the measure to establish a consistent approach to highly effective age assurance in the Illegal Content Codes, alongside minor amendments to the Part 3 HEAA Guidance.

5.62 By aligning the definition of highly effective age assurance across the Illegal Content and Protection of Children Codes and setting out clear principles for its implementation, the measure provides regulatory coherence and consistency for providers and supports the effective application of child-specific safety measures where age assurance is used. The accompanying updates to the Part 3 HEAA Guidance ensure that Part 3 services can refer to a single source of guidance, reducing regulatory complexity.

5.63 Taken together, these changes strengthen the operation of the Illegal Content and Protection of Children Codes of Practice for user-to-user services by supporting consistency in the regulatory framework, alongside accuracy and proportionality in implementation, without mandating specific technologies or altering the scope of existing duties.

Measures on appropriate action for age assessment appeals

5.64 In our June 2025 Consultation, we proposed to introduce measures concerning how providers should deal with age assessment appeals.

5.65 In our proposals, we acknowledged that while we consider that the use of highly effective age assurance will lead to the effective application of safety measures for children, the age

[387] Ofcom and ICO (2026), Age Assurance: A Joint Statement by Ofcom and the Information Commission's Office.

[388] See the Joint Statement by Ofcom and the Information Commission's Office on Age Assurance for further details on how service providers can meet their obligations under both the Act and UK data protection legislation when implementing age assurance.

of some users may be incorrectly assessed. This may have an impact on the online experience and rights of a user whose age is incorrectly assessed, as well as on other users.

Our decision

5.66 We have decided to proceed with the measures as proposed in our June 2025 Consultation. We set out our reasoning in the following sections.

5.67 The full text of the measures can be found in the Illegal Content user-to-user Codes of Practice and the measures are referred to as ICU D15 and D16.

How these measures work

5.68 We have amended the Illegal Content Codes to the appropriate action that should be a taken where user considers their age or age range has been incorrectly assessed.

5.69 We are introducing two measures:

a) ICU D15 which applies to large or multi-risk services using highly effective age assurance; and
b) ICU D16 which applies to other services (that are neither large nor multi-risk) using highly effective age assurance.

5.70 We are defining "age assessment appeals" as a complaint by a UK user whose experience on the service is impacted because measures used to comply with illegal content safety duties (for example, the grooming measures referred to in paragraph 5.3) have resulted in an incorrect assessment of the user's age or age range.[389]

Appropriate action for age assessment appeals (services that are large or multi-risk)

5.71 We have decided to add a provision to the Illegal Content Codes stating that service providers which are large or multi-risk should:

a) Have regard to the following matters in determining what priority to give to consideration of an age assessment appeal:
    i) the seriousness of the impact on the user as a result of the assessment of their age;
    ii) whether the decision made on the basis of the user's age was made without human oversight and, if so, information available about the accuracy of the specific technology used in making age assessments of the type concerned;
    iii) the past error rate on the service in relation to age assessments of the type concerned; and
    iv) any representations made by the user as part of the complaint as to the effect of the decision on their livelihood.
b) Monitor their performance against performance targets relating to:
    i) the time it takes to determine the age assessment appeal; and
    ii) the accuracy of decision making.
c) Resource the service to give effect to those targets.

[389] As a consequence of introducing the defined term "age assessment appeal", we are also changing the existing defined term "appeal" to "content appeal". This is necessary to distinguish it from an "age assessment appeal".

d) Act if the provider determines that the user's age was incorrectly assessed by taking any necessary steps to restore the user to the position they would have been in had the assessment been correct, as far as is appropriate and possible.
e) Monitor trends in age assessment appeals to help improve any age assurance process used on the service.

Appropriate action for age assessment appeals (services that are neither large nor multi-risk)

5.72 We have decided that that service providers which are neither large nor multi-risk should:

a) determine age assessment appeals promptly.
b) act if the provider determines that the user's age was incorrectly assessed by taking any necessary steps to restore the user to the position they would have been in had the assessment been correct, as far as is appropriate and possible.
c) monitor trends in age assessment appeals to help improve any age assurance process used on the service.

Differences from equivalent measures in the Protection of Children user-to-user Code

5.73 These measures are similar to measures PCU D11 and D12 in the Protection of Children Code. The key difference is that in PCU D11 and D12 the definition of "age assessment appeals" and the description of the action that the provider must take if the appeal is upheld refer to restricting access to content.

5.74 However, the measures in this statement that use highly effective age assurance are not measures that restrict access to content, instead they involve:

a) Restriction of functionalities (for example, ICU E2 and ICU E3, user-to-user service measures where a user that has not been determined to be an adult is not recommended potentially illegal content).
b) Settings being off by default (for example, measure ICU F1.6 states that where a service displays automated location information, it should implement default settings on the account of a user that has not been determined to be an adult ensuring that the location information associated with that account is not visible to other users of the service).
c) The provision of supportive information in certain circumstances, including when a user that has not been determined to be an adult seeks to change certain default setting.

5.75 As such, it would not be appropriate to refer to restricting access to content in the definition of "age assessment appeals" and in the description of the action that the provider must take if the appeal is upheld. Therefore, we have:

a) defined "age assessment appeals" by amending reference from a "user who is unable to access content" to "a user whose experience on the service is impacted because measures used to comply with illegal content safety duties have resulted in an incorrect assessment of the user's age"; and
b) described the action that the provider must take if the appeal is upheld as taking any necessary steps to restore the user to the position they would have been in had the assessment been correct, as far as appropriate and possible.

Our reasoning

Effectiveness and benefits

5.76 These measures ensure there is an effective appeal process when users have been subjected to an incorrect age assessment, which may result in their access to parts of the service or functionalities being constrained. An appeals mechanism restores the user – as far as is possible and appropriate – to the position they would have been in if the provider determined their age correctly.

5.77 Service providers monitoring trends in complaints about incorrect assessments of a user's age will help to ensure the age assurance process is working effectively.

5.78 To the extent that we are introducing additional provisions for large or multi-risk services, these relate to the prioritisation of age assessment appeals. Such services are likely to receive a high volume of appeals, given the greater number of users undergoing initial age assessments and would therefore benefit from prioritisation processes and performance targets to manage these appeals effectively.

5.79 A range of respondents across civil society, government and industry emphasised that age assessment appeals are an important safeguard, which must be designed in a way that does not undermine the effectiveness or integrity of age assurance processes.[390] They noted that age assurance systems are not infallible and that users should have access to clear and effective routes to challenge incorrect assessments, and highlighted the importance of timely processes and protections against bias or inaccurate assessments, while ensuring appeals remain consistent with data protection principles.[391]

5.80 The National Society for the Prevention of Cruelty to Children (NSPCC) specifically welcomed the introduction of age assessment appeals but warned that the cost of handling them may incentivise services to overestimate ages in order to limit both the volume of appeals and the financial burden. To avoid this, it recommended benchmarks for technical accuracy and dedicated reporting channels for users to flag where they think a user account has incorrect age.[392]

5.81 Several stakeholders raised concerns about appeals processes, including biometric technologies not supporting user specific appeals without privacy risks, limited appeal opportunities when using reusable tokens and the need to balance user-redress with security and privacy.[393]

5.82 We recognise concerns that appeals must not undermine the integrity of highly effective age assurance. While services must provide routes for user redress, these cannot weaken or bypass highly effective age assurance, nor justify practices such as systematically

---

[390] National Society for the prevention of Cruelty to Children (NSPCC) response to the June 2025 Measures Consultation, p.46; NCMEC response to the June 2025 Consultation, p.7 ODDA response to the June 2025 Consultation, p.8; Online Responsibility Network response to the June 2025 Consultation, p. 4; [✂].
[391] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.21; Digital Resilience in Education - Welsh Government response to the June 2025 Consultation, p.14; Evangelical Alliance response to the June 2025 Consultation, p.9; ICO response to the June 2025 Consultation, p.30; Internet Matters response to the June 2025 Consultation, pp.16-17; Open Rights Group response to the June 2025 Consultation, p.38; Verifymy response to the June 2025 Consultation, pp.9-10; [✂].
[392] NSPCC response to the June 2025 Consultation, p.46-48.
[393] ODDA response to the June 2025 Consultation, p8; Online Responsibility Network response to the June 2025 Consultation, p. 4; [✂].

overestimating ages to reduce appeal volumes. In particular, providers should address accuracy issues, including any bias toward treating users as adults, and take corrective action where identified. This approach supports the effective operation of the measure by enabling correction of errors while preserving the integrity of highly effective age assurance.

5.83 We also note concerns about privacy, biometrics, and limitations of reusable tokens. Our approach does not mandate any specific technology, and appeals processes must remain secure, proportionate, and privacy-preserving, with clear ways for users to report incorrect age assessments. Taken together, these expectations ensure appeals are effective, efficient, and safeguarded against misuse, while encouraging improvements in initial accuracy.

5.84 Having considered the June 2025 Consultation responses on appeals, we are maintaining our position as proposed, as the feedback does not provide evidence that would warrant changes to our approach.

Impact and costs

5.85 We have considered impacts on service providers who use highly effective age assurance for the purpose of targeting measures included in the Illegal Content Codes. We note that all providers within the scope of these measures will also be within the scope of equivalent measures concerning appeals against action taken by providers against suspected illegal content.[394]

5.86 As the activities and costs involved in the measures are likely to be similar, we expect there to be some overlap in actions required and the associated costs. Despite this, there is likely to be some increase in costs to implement these measures. This could be due to the increased number of appeals, additional costs associated with appeals related to age assurance (for example, hiring staff with expertise on age assurance), and monitoring the performance of age assurance decisions.

5.87 We would generally expect the volume of appeals a service provider receives to increase with the size of the service because larger services are likely to make a greater number of age assessment decisions. This means that the highest costs will likely be incurred by the service providers of the largest services who are most likely to be able to absorb them, and who we expect would deliver the greatest benefits from implementing these measures.

5.88 We are introducing additional prioritisation measures for large or multi-risk services. As noted in paragraph 5.78, we consider these will aid their effective handling of age assessment appeals. These services are also already subject to a similar measure related to appeals (see measure ICU D8). We consider the incremental cost associated with the equivalent age assessment appeal prioritisation process is likely to be limited.

5.89 Overall, we consider the additional costs identified in paragraph 5.86 are likely to be small, relative to the benefits these measures provide in increasing user satisfaction and supporting the effective operation of age assurance processes.

[394] Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes), Measures ICU D8-ICU D10A.

Rights impact

Freedom of expression and association

5.90 Article 10 of the European Convention on Human Rights (ECHR) sets out the right to freedom of expression, which encompasses the right to hold opinions and to receive and impart information and ideas without unnecessary interference by a public authority. Article 11 sets out the right to associate with others. We must exercise our duties under the Act in light of users' and services' Article 10 and 11 rights and not interfere with these rights unless we are satisfied that to do so is prescribed by law, pursues a legitimate aim, is proportionate to the legitimate aim, and corresponds to a pressing social need.

5.91 We did not receive consultation responses identifying any distinct or additional issues which impact freedom of expression and association, as a result of the proposed measures on age assessment appeals.

5.92 We do not consider service providers taking appropriate action in response to age assessment appeals will interfere with individuals' rights to freedom of expression and association. We determine that the measures will have positive impacts on the freedom of expression and association, because the measures are aimed at providers remedying restrictions applied to users' accounts.

Privacy and data protection

5.93 Article 8 of the ECHR sets out the right to respect an individual's private and family life. We must exercise our duties under the Act in light of users' privacy rights and not interfere with these rights unless we are satisfied that to do so is prescribed by law, pursues a legitimate aim, is proportionate to the legitimate aim, and corresponds to a pressing social need.

5.94 The ICO supported the inclusion of appeals processes for age-assurance assessments in the Illegal Content Codes and noted that such processes are an important safeguard for data protection rights, helping services uphold fairness and accuracy principles and users' right to rectification.[395] It stressed, however, that organisations implementing highly effective age assurance must continue to meet all data protection obligations –such as taking reasonable steps to correct or erase inaccurate personal data – and that considering Ofcom's guidance or Codes does not, in itself, guarantee compliance with data protection law. The ICO therefore requested a clearer distinction between following Ofcom's guidance and complying with data-protection legislation, which remains solely within the ICO's remit.[396]

5.95 Two civil society organisations, one academic association, three individual respondents and two industry stakeholders said that appeals processes must minimise personal data collection and include strong safeguards to prevent privacy and security risks.[397] They raised concerns about the increased collection of sensitive information, unclear protections against breaches or misuse, the need to limit retention and guard against malicious or repeated appeals, and the broader risks of identity theft or blackmail if verification data were compromised.

5.96 We welcome the ICO's support and reiterate our statements in the Part 3 HEAA Guidance and January 2025 Statement that all age assurance methods involve the processing of

[395] ICO response to the June 2025 Consultation, p.29.
[396] ICO response to the June 2025 Consultation, p.29.
[397] [✂]; [✂]; Open Rights Group response to the June 2025 Consultation, p.37-38; [✂]; [✂].

personal data, and as such, service providers who implement them are subject to the requirements of the UK's data protection regime, which is overseen and enforced by the ICO, in addition to their duties under the Act.

5.97 We decided that the measures overall will have a positive impact on privacy rights, due to the measures increasing transparency and accountability around how service providers make age assessments. Additionally, the measures will incentivise service providers to reduce the number of incorrect age assessments, so they may save resources around complaint handling.

5.98 We also decided that our measures will support providers to develop a clear mechanism through which individuals can safeguard the accuracy of their personal data held by the provider. We highlighted the importance of providers' compliance with data protection law and to familiarise themselves with relevant ICO guidance when developing age assessment appeals processes.

Who these measures apply to

5.99 In the June 2025 Consultation, we said that the measures should apply to services that use highly effective age assurance to determine which UK users of the service are not determined to be adults for the purpose of targeting measures in the Illegal Content Codes at such users, their user accounts, or their content feeds. This is because the risk that a user may be impacted by an incorrect assessment of age or age range will only arise in this circumstance.

5.100 We are therefore recommending these measures for services that use highly effective age assurance to determine which UK users are not determined to be adults for the purpose of targeting measures in the Illegal Content Codes.

5.101 Having considered consultation responses on scope, we are confirming this approach and are not making changes.

Conclusion

5.102 We have decided to proceed with the measures on appropriate action for age assessment appeals as proposed in the June 2025 Consultation. The measures establish clear and proportionate expectations for how providers should handle appeals where an incorrect age assessment has affected a user's experience, ensuring that users have access to meaningful redress while preserving the integrity and effectiveness of highly effective age assurance.

5.103 The measures provide flexibility for providers in how appeals processes are designed and operated, recognise differences in scale and risk between services, and include additional expectations for large or multi-risk services where appropriate. Taken together, they support accurate age assurance outcomes, help prevent systemic errors or unintended outcomes, and ensure appeals function as a safeguard without weakening age assurance processes.

5.104 Having considered consultation responses, we are satisfied that the measures strike an appropriate balance between user redress, proportionality, and the effective operation of the Illegal Content Codes.

5.105 The full text of the measures can be found in the Illegal Content Codes of Practice for user-to-user services and the measures are referred to as ICU D15 and D16.

Overall conclusion

5.106 Having considered the full range of consultation evidence, we are confirming the measures as proposed. We consider that the benefits of aligning the definition of highly effective age assurance across the Codes of Practice, introducing accompanying principles, and a proportionate appeals approach to the Illegal Content Codes, outweigh the associated costs and risks.

5.107 These measures provide clear advantages, including:

- strengthening child protection by enabling safeguards to be effectively targeted;
- increasing consistency across the Illegal Content and Protection of Children Codes;
- improving the accuracy, fairness, and transparency of age-assurance processes; and
- supporting user redress without weakening the integrity of the underlying systems.

5.108 During the June 2025 Consultation, stakeholders highlighted cost, privacy, operational and circumvention risks related to the measures. We consider these risks are mitigated through a proportionate, flexible approach that limits costs for smaller services, and supports privacy preserving implementation to safeguard users' rights. Evidence from the Age Assurance Report also indicates that highly effective approaches are significantly more resilient to circumvention.

5.109 Balancing these benefits against the costs, risks, and rights impacts, we conclude that the measures are proportionate and appropriate for achieving the statutory objectives of reducing children's exposure to illegal harms while ensuring protections remain targeted, evidence-based and future-proofed.

6. Increasing the effectiveness of safety defaults and supportive messaging

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

What is this chapter about?

Our Codes state that, in order to protect children from grooming, children's accounts should have certain risky functionalities turned off by default. These functionalities include network expansion prompts, connection lists, location displays and certain direct messaging. In the June 2025 Additional Safety Measures Consultation, we proposed to amend the Codes to specify that, for the purposes of implementing these protections against grooming, service providers should either use highly effective age assurance or apply the measure to all users. This chapter explains our rationale to proceed with this amendment.

What decisions have we made?

We have decided to proceed with the amendments. Our final Code measures are summarised in the table below:

Number in Codes	Recommended measure	Who should implement this
ICU F1	When implementing safety defaults and supportive messaging as recommended by ICU F1 and ICU F2, providers should either: i) use highly effective age assurance, as set out in the Part 3 HEAA Guidance to target the measures at child users; or ii) apply ICU F1 and F2 to all users of the service	This is an amendment to measures ICU F1 and F2, therefore the measure will apply to: a) regulated user-to-user services at high risk of grooming; and/or b) regulated user-to-user services that are at a medium risk of grooming, and are a large service (7 million or more monthly UK users).
ICU F2		

Why have we made these decisions?

Highly effective age assurance is a better way of determining which users are not determined to be adults than other methods such as self-declaration of age. This amendment to the Codes should therefore increase the proportion of children who benefit from protections against grooming.

Introduction

6.1 In this chapter, we set out our decision to amend the existing measures on safety defaults and support for child users included in the Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes) to specify that service providers should either i) use highly effective age assurance to determine which users to apply the measures to, or ii) apply the

measures to all users by default. We are updating the measures on how they are implemented to reflect our approach to highly effective age assurance, as set out in our draft Part 3 Guidance on Highly Effective Age Assurance.[398] [399] These measures are designed to combat grooming for the purpose of child sexual exploitation and abuse (CSEA) on user-to-user services.

Our approach

Our approach to the measures in the Illegal Content Codes

6.2 Our December 2024 Statement on Protecting People from Illegal Harms (December 2024 Statement) set out two measures that service providers should take to protect children from grooming.

a) The first measure on safety defaults makes it harder for perpetrators to find and interact with children by providing that:[400]

   i) child users are removed from, and not presented with, network expansion prompts;
   ii) child users are not included in connection lists;
   iii) users are not able to direct message child users without an established connection or child users are provided with a means of actively confirming to receive direct messages if there is no user connection functionality; and
   iv) automatic location information display is turned off.

b) The second measure provides that child users receive timely supportive information at four key user journey points:[401]

   i) when children are turning off defaults;
   ii) establishing a new connection with a user;
   iii) sending a direct message to a user for the first time; and
   iv) taking action against an account.

Our approach to increasing the effectiveness of the measures

6.3 Our December 2024 Statement predated the publication of the final guidance on highly effective age assurance for providers of Part 3 services (Part 3 HEAA Guidance). This is our guidance to assist providers of regulated user-to-user and search services (Part 3 services) in implementing highly effective age assurance for the purpose of fulfilling their regulatory obligations under the Act. We decided to implement measures to protect children from grooming despite the lack of final guidance at the time, in order to bring in the important protections they would afford to users.

6.4 Therefore, we recommended that providers should apply the measures "to the extent that it has an existing means to determine the age or age range of a particular user of the

[398] Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes), Measures ICU F1 and ICU F2.
[399] The aim of online grooming is to manipulate or coerce children into engaging in sexual activity. Grooming offences can include, but are not limited to, sexual or indecent communications with children, engaging in sexual activity with a child, or in the presence of a child, or coercing or inciting the child into sexual activity.
[400] Illegal Content Codes, Measure ICU F1.
[401] Illegal Content Codes, Measure ICU F2.

service". We stated that service providers could continue to use their "existing means" to apply the measures.

6.5 Having now finalised our position on highly effective age assurance, our proposed amendments in the June 2025 Additional Safety Measures Consultation (June 2025 Consultation) recommended that providers with services in scope of the measures should implement the measures using one of the following approaches:

a) Option A**:** Using highly effective age assurance, as set out in the Part 3 HEAA Guidance, to target those not determined to be an adult; or
b) Option B: Applying the two measures to all users of the service.

Our decision

6.6 Overall, we have decided to confirm the measures as proposed in our June 2025 Consultation and outline our reasons for doing so in the following sections. We have made some minor changes to ICU F1.5 to clarify the definition of a 'specified connection'. This amendment does not alter the measures, but makes the definition simpler, to ensure the recommendation are consistent and clear for providers to consider and implement where applicable.

6.7 The full text of the measures can be found in our Illegal Content Codes and are referred to as ICU F1 and F2. The measures are part of our Code of Practice on child sexual exploitation and abuse (CSEA) duties.

Our reasoning

Effectiveness and benefits

6.8 Grooming remains a prevalent online harm. Grooming is thought to be significantly under-reported and although it is not possible to accurately determine the full scale of online grooming, the available evidence indicates it is a significant and growing harm. New reporting requirements have shown that in 2025, the National Center for Missing & Exploited Children (NCMEC) received 1.4 million reports of online enticement, a 156% increase from 2024.[402] [403]

6.9 Evidence from law enforcement, child protection organisations and stakeholder responses to the June 2025 Consultation indicates that online grooming for the purpose of child sexual exploitation and abuse continues to occur at scale[404] across services frequently used by children.[405]

---

[402] Online enticement, as defined by the National Center for Missing and Exploited Children (NCMEC) is a crime which involves an adult communicating with a child for sexual purposes and includes sextortion. This is a broad category of online exploitation and includes sextortion, in which a child is being groomed to take sexually explicit images and/or ultimately meet face-to-face with someone for sexual purposes, or to engage in a sexual conversation online or, in some instances, to sell/trade the child's sexual images. This type of victimization takes place across every platform; social media, messaging apps, gaming platforms, etc.
[403] National Center for Missing & Exploited Children (NCMEC) 2025. Annual Report.
[404] Internet Watch Foundation (IWF) response to the June 2025 Consultation, p.52.
[405] Ofcom (2024), Illegal Harms Register of Risks (Illegal Harms Register), chapter 2A: Grooming; Institution of Engineering and Technology response to the June 2025 Consultation, p.2; Internet Matters response to the June 2025 Additional Safety Measures Consultation (June 2025 Consultation), p.19.

6.10 The impact of grooming on children is significant and can be long lasting. While the effects vary between victims and survivors, it can result in sustained psychological and emotional harm. Further detail on the nature and impact of grooming is set out in the Illegal Harms Register of Risks (Illegal Harms Register) and in our December 2024 Statement.

6.11 Our analysis shows that where the measures in the December 2024 Statement are applied to children's accounts, they will materially reduce the risk of those children being groomed online by perpetrators.[406]

6.12 While the measures in the December 2024 Statement have the potential to reduce the risk of grooming, it also left open the possibility of service providers relying solely on a user's self-declaration of age to determine which users were children. The December 2024 Statement also stipulated that service providers only needed to apply the measures to the extent that they had an existing means of determining which of their users were children.

6.13 Therefore, significant numbers of children may not, until now, have benefitted from the package of measures introduced in December 2024 – either because some could easily bypass protections by misrepresenting their age or because they were using a service which did not have a means to determine age.

6.14 The decision we are taking in this statement closes these gaps. In our June 2025 Consultation, we explained that by amending the Illegal Content Codes to stipulate that service providers should either use highly effective age assurance or apply the measures to all users, we would materially increase the number of children that benefit from the protections of the measures.

6.15 In addition, more services at risk of grooming will be expected to implement the measures on safety defaults and support for child users, because not having an existing means to determine a user's age will no longer exclude a service from implementing the measures.[407] In the June 2025 Consultation we therefore explained that we estimated the benefits of this measure would be significant.

6.16 Numerous responses to the June 2025 Consultation endorsed our assessment and highlighted that highly effective age assurance can help to reduce risks of grooming.[408]

6.17 Some stakeholders raised concerns about the consequences of giving service providers the option of applying the measures to all users rather than using highly effective age assurance to identify which users were children.[409] However, these stakeholders did not provide evidence to explain why our proposed approach would not be effective at protecting children from grooming, and it therefore did not give us grounds to revise the assessment of the efficacy and benefit of the measure from the June 2025 Consultation.

6.18 We therefore conclude that our approach to increase the effectiveness of the measures will deliver significant benefits to users.

---

[406] Ofcom (2024), December 2024 Statement, Vol.2, chapter 8: U2U settings, functionalities and user support.
[407] The measures apply to a provider in respect of each service it provides that is either i) at high risk of grooming or ii) is a large service that is at medium risk of grooming.
[408] Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p.4; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.15; [✂]; Internet Matters response to the June 2025 Consultation, p.18; IWF response to the June 2025 Consultation, p.51; Marie Collins Foundation response to the June 2025 Consultation, p.11.
[409] Marie Collins Foundation response to the June 2025 Consultation, p.11; National Society for the Prevention of Cruelty to Children (NSPCC) response to the June 2025 Consultation, p.49.

Impacts and costs

Option A: Costs of applying highly effective age assurance and safety defaults

Direct costs

6.19 The approach to implementing the measures in line with Option A would give rise to two types of direct cost: (i) cost of applying highly effective age assurance; and (ii) the cost of applying the safety defaults.

Costs of applying highly effective age assurance

6.20 Some services in scope of these measures are already required to use highly effective age assurance to comply with their age assurance duties under the Act and/or through implementing the Protection of Children Code of Practice for user-to-user services. We expect that they will be able to apply existing highly effective age assurance processes when implementing these measures and therefore will incur relatively small additional costs. Service providers not currently using highly effective age assurance would incur far more significant costs.

6.21 In the June 2025 Consultation, we estimated the costs of implementing highly effective age assurance using third-party provision or by developing an in-house solution. We have updated our estimates of the cost of in-house provision to reflect wage-based inflation; however, these estimates remain unchanged due to rounding.[410] We did not receive feedback which contradicted our cost estimates for third-party provision or in-house development. We outline the costs for these two means of provision below.

Cost of third-party provision

6.22 We maintain our cost estimates from the June 2025 Consultation:[411]

- there is a fixed cost per check of between £0.05 and £0.30;
- each existing user is checked once leading to a one-off cost;
- the cost per check will remain unchanged over time; and
- each new user will be checked once and these ongoing age checks on new users will continue annually.

6.23 In addition to the cost of age checks, there may be upfront costs linked to (1) the age assurance provider setting up a client account to prepare the age assurance method for use; (2) the service provider having to introduce changes to the existing ICT infrastructure or building of a new user interface to integrate age assurance with the service; and (3) the service having to train some of its staff when the process becomes operational.

6.24 We did not receive feedback in response to our June 2025 Consultation challenging these cost estimates. Some stakeholders referenced alternative cost estimates for highly effective age assurance which appear to fall within our cost estimate range.[412]

---

[410] Refer to annex 5: Further detail on economic assumptions and analysis for a breakdown of economic assumptions and costing methodology.
[411] See Table A5.2 in annex 5: Further detail on economic assumptions and analysis for further detail on the cost of age checks using third party providers.
[412] Edwards, L. response to the June 2025 Consultation, p.3; [✂].

6.25 In our recently published report on the use of age assurance, we presented analysis on the costs of implementing highly effective age assurance.[413] Based on information collected from a sample of regulated services, we maintained that third-party per-check costs for the services in our sample remain broadly consistent with the assumptions used in our June 2025 Consultation. While this information was primarily obtained from larger services, the reported costs were generally towards the lower end of our estimated range. Taken together, these findings provide support for our estimated range of third-party per-check HEAA costs.[414]

Cost of in-house highly effective age assurance provision

6.26 In our June 2025 Consultation, we estimated that there would also be ongoing staff costs relating to monitoring, evaluation and maintenance, and there could be additional ongoing data costs if the method requires significant improvements and/or changes in the future. We maintain that these ongoing staff costs could reach £1 million annually or potentially more, depending on a service's approach. Our estimates are based on the same salary assumptions for upfront and ongoing costs.[415]

Cost of applying safety defaults

6.27 We expect that most providers will have previously incurred the cost of implementing safety defaults and supportive information for child users. We have updated our cost estimates since our December 2024 Statement using the latest salary data from the ONS.

6.28 We estimate that a relevant service provider will incur one-off costs of approximately £11,000 to £359,000 to implement safety defaults and annual maintenance costs of £3,000 to £90,000.[416] To implement supportive information, we estimate a provider would incur one-off costs of £32,000 to £359,000 and annual maintenance costs of £8,000 to £90,000.[417] Smaller providers will incur costs nearer the low estimates as they will be less likely to need to make significant system changes.

6.29 Those service providers which have not already implemented these measures, as they did not have an existing means to determine the age of users, will incur these costs, in addition to the cost of using highly effective age assurance. Where services have parts of these measures in place already, costs will be lower for their providers.

Indirect costs

6.30 The June 2025 Consultation discussed potential indirect costs associated with services losing traffic (if users do not want to go through highly effective age assurance), as well as

---

[413] Ofcom (2026), Report on the use of age assurance, p. 49-52. This sample is not representative. Where we cite statistics or findings related to analysed services, they apply only to the services in the sample and should not be interpreted as indicative of sector-wide practices.
[414] See annex 5: Further detail on economic assumptions and analysis for further detail on the evidence collected from services on third-party per-check costs in the context of our cost estimates.
[415] Cost estimates based on the assumption that software engineers equivalent to 14 FTE will be required for system maintenance annually. See annex 5 for further details on our assumptions for age assurance costs.
[416] For safety defaults, we estimate it would take one to six months of software engineering time matched with equal time from professional occupation staff, with an additional zero to twenty-four months of time from professional occupation staff to conduct front and back end system changes.
[417] For supportive information, we estimate it would take three to six months of software engineering time matched with equal time from professional occupation staff, with an additional zero to twenty-four months of time from professional occupation staff to conduct system changes. See annex 5: Further detail on economic assumptions and analysis.

adult users potentially experiencing friction and impacted service functionalities from highly effective age assurance implementation costs.

6.31 Two stakeholders raised concerns that requiring highly effective age assurance could impose significant costs on small services and harming overall UK innovation in the digital sector.[418] This measure also only applies to small services where there is a high risk of grooming. Therefore, any impact on UK innovation that arises from the costs of the measure is likely to be proportionate when seen in the context of the benefits that arise from reducing the potential for grooming on a high-risk service.

Option B: Costs for applying the measures to all users

Direct costs

6.32 Our June 2025 Consultation stated that service providers who apply Option B could avoid incurring the costs of implementing highly effective age assurance (where they are not already required to use highly effective age assurance to comply with other duties). Ukie stated Option B would be less costly overall and more in line with frameworks currently used to moderate content.[419]

6.33 We have outlined our updated costs of implementing the safety defaults and supportive information measures for children in paragraph 6.28. For in-scope providers which have already incurred this cost of implementation for child users, we consider that applying these measures to all users has negligible incremental direct cost.

6.34 Providers who did not previously implement the measures because they did not have an existing means to determine the age of users will incur the costs described in full.

Indirect costs

6.35 Service providers may incur indirect costs associated with implementing the safety defaults and supportive information. Limiting interactions between users may lead to users spending less time on a service which may impact subscription or advertising revenues. However, because these settings are defaults and can be turned off, we consider these effects to be small for most types of services. Supportive information may also have indirect impacts on user experience, but this is mitigated by the fact that providers have the flexibility to design the information to reduce unnecessary interference. Our December 2024 Statement outlined these impacts in more detail.[420]

6.36 Having considered responses to our impact assessment and feedback on the costs of the options, we are confirming the cost estimates as stated in our June 2025 Consultation.

Impact on users

Wider impacts on child users

6.37 Alongside the benefits in reducing risk of grooming, the measures may have some smaller adverse impacts on how children interact online, including affecting their ability to make new friends online. Less frequently, it could also affect the ability of children to monetise their online presence by attracting users to their content. However, child users who feel that their online experience is strongly disrupted can turn off the safety defaults if they wish to do so. While there are possible adverse effects, limiting children's interactions

[418] Association for UK Interactive Entertainment (Ukie) response to the June 2025 Consultation, p.38; [✂].
[419] Ukie response to the June 2025 Consultation, p.38.
[420] Ofcom (2024), December 2024 Statement, Vol.2, chapter 8: U2U settings, functionalities, and user support.

primarily helps mitigate their identification and coercion by perpetrators. Our December 2024 Statement provides additional evidence on how this is achieved.[421] Therefore, the net effect the measure will have on children is overwhelmingly positive.

Wider impacts on adult users

Option A (using highly effective age assurance)

6.38 There may be an inconvenience to some adult users of going through the highly effective age assurance process if a service provider adopts Option A. The adult user may be concerned about providing their personal data to demonstrate they are an adult. However, we consider it most likely that a service provider may provide the option to adult users of going through the highly effective age assurance process or having children's settings applied to their account. In this case the adverse effect on adult users will be those associated with Option B and are described in the following paragraphs. Where an adult user does go through the age assurance process, there would be processing of personal data. We address this in the 'Privacy' section.

Option B (applying to all users)

6.39 Adult users could be adversely affected by the inconvenience of having to manually turn off the safety defaults applied to their account, if they want to, and having supportive messaging prompts applied to their account that cannot be turned off.[422]

6.40 Our June 2025 Consultation set out that having fewer connection suggestions might have a negative impact on adult users. Snap Inc. said it introduced network expansion prompts to prevent users sharing usernames via more dangerous means (such as public forums).[423] One provider submitted that applying these settings to all users would make it harder for users to find people they already know.[424]

6.41 The Illegal Harms Register explained network expansion prompts can play a major role in facilitating child grooming. Therefore, we see the negative impacts on adult users as being outweighed by the benefit of reducing the risk of harm of online grooming. Users who want to change the default settings will be able to do so manually, which reduces the adverse impact on adult users. Some adult users may also experience increased privacy benefits from the measures.

Rights impact

6.42 Our decision to increase the effectiveness of the safety defaults and support to child users measures may impact users' rights to freedom of expression and association and have implications on their privacy rights. However, we consider that the degree of interference with these rights will be subject to the option taken by the provider to either apply the measures to all users, or to target which users they apply the measures to by using highly effective age assurance. Where highly effective age assurance is used, we consider that the type of age assurance method(s) used will be a factor that impacts the degree of interference with users' rights.

---

[421] Ofcom (2024), December 2024 Statement, Vol.2, chapter 8: U2U settings, functionalities, and user support.
[422] There is a possible scenario where a service has highly effective age assurance and uses it to allow adult users to turn off supportive messaging but applies safety defaults to all users.
[423] Snap Inc. response to the June 2025 Consultation, p.17.
[424] [✂].

Freedom of expression

6.43 Article 10 of the European Convention on Human Rights (ECHR) sets out the right to freedom of expression, which encompasses the right to hold opinions and to receive and impart information and ideas without unnecessary interference by a public authority. Article 11 sets out the right to associate with others. We must exercise our duties under the Act in light of users' and services' Article 10 and 11 rights and not interfere with these rights unless we are satisfied that to do so is prescribed by law, pursues a legitimate aim, is proportionate to the legitimate aim, and corresponds to a pressing social need.

6.44 The feedback we received from stakeholders to our initial rights assessment in the June 2025 Consultation highlight issues that we previously addressed in the December 2024 Statement and the June 2025 Consultation. We acknowledge this feedback and have reiterated our positions in response to the feedback in annex 1: Further stakeholder responses. We set out our final rights assessment below.

Final rights assessment

6.45 Overall, we determine that our decision to increase the effectiveness of safety defaults and address the severe risk of grooming harms to children of all age groups is proportionate to the impact on users' rights of freedom of expression and association.

6.46 Our decision removes the application requirement of "existing means" from the safety default and supportive for child user measures. This amendment broadens the scope of both measures, which will result in the measures impacting more children.

6.47 In the December 2024 Statement, we set out that the impact of these measures on users' rights to freedom of expression would be that they were unable to form connections and engage through the relevant functionalities.[425] However, we noted that this impact would be negligible as the safety defaults could be manually turned off and the support for child user measures has a negligible impact on users' rights.

6.48 We expect our decision to make more children subject to the safety defaults and consequently that more child users' rights will be impacted as described in the December 2024 Statement. However, we consider that the increase in the number of children affected by the measures does not change the proportionality of the impact described in the December 2024 Statement. We remain of the view that the overall impact on child users' rights to be proportionate to address the risk of grooming on services.

6.49 Our decision may impact adult users' rights where highly effective age assurance is implemented. Highly effective age assurance could increase friction in the user experience, which could lead to adults choosing not to use a service and consequently restrict their ability to send and receive information through a service. However, we consider this impact to be limited, as it is ultimately the adult user's choice as to whether they use a service and it will be in providers' interests to implement age assurance in a manner which will minimise user experience friction for its users.

6.50 We acknowledge the impact on adult users' rights, if incorrect assessments occurred on services where highly effective age assurance is implemented. However, where incorrect assessments occur and adults are assessed to be under 18, the impact will not be disproportionate, as adult users will be able to turn off the safety defaults and the impact of the supportive information is negligible. Furthermore, where providers implement age

[425] Ofcom (2024), December 2024 Statement, Vol.2, chapter 8: U2U settings, functionalities, and user support.

assurance in line with our recommendations and Part 3 HEAA Guidance[426], we expect there to be limited incidences of age checks resulting in an adult user being wrongly determined as a child, which should further limit any impact on adult users' rights.

6.51 Where the safety default and support for child users' measures are applied to all users the impact on adult users' rights will be negligible because adult users will be able to turn off the safety defaults and the impact of the supportive information is negligible.

Privacy

6.52 Article 8 of the ECHR sets out the right to respect an individual's private and family life. We must exercise our duties under the Act in light of users' privacy rights and not interfere with these rights unless we are satisfied that to do so is prescribed by law, pursues a legitimate aim, is proportionate to the legitimate aim, and corresponds to a pressing social need.

6.53 We received support from the ICO in response to our rights assessment on privacy in our June 2025 Consultation, particularly our recommendations to ensure compliance with data protection law requirements.[427] Additional feedback on data privacy and security was received from other stakeholders that we have previously addressed in the January 2025 Statement and the April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement). We summarise this feedback in annex 1: Further stakeholder responses and reiterate our positions from the previous statements.

Final rights assessment

6.54 All methods of age assurance involve processing some personal data of individuals. Therefore, where highly effective age assurance is implemented, it will have an impact on users' rights to privacy, including their rights under data protection law. However, the degree of interference with users' privacy rights will depend on the age assurance process that the provider intends to implement, and the nature of the information needed by the age assurance process. Furthermore, we do not expect providers who apply the measures to all users to collect personal data or information about users more than they already do.

6.55 We expect any impact on rights to privacy to be proportionate, because providers that implement safety measures, including highly effective age assurance, are expected to comply with relevant data protection legislation when processing personal data and have a duty to have regard to the importance of protecting users from breach of rules related to privacy rights.[428]

6.56 The Part 3 HEAA Guidance outlines relevant considerations for providers when dealing with privacy rights and implementing age assurance. This includes relevant ICO Guidance and data protection legislation, which explains the most relevant rules and principles which apply to age assurance.[429] The Part 3 HEAA Guidance also gives examples, on how providers can demonstrate that they have had regard to the importance of protecting users from a breach of their privacy and data protection rights.[430]

[426] Ofcom (2025), Guidance on highly effective age assurance.
[427] ICO response to the June 2025 Consultation, p.29.
[428] Section 22(3) of the Online Safety Act 2023.
[429] See 5.3-5.10 of Part 3 HEAA Guidance.
[430] See 5.11-5.13 of Part 3 HEAA Guidance.

6.57 As set out in our Part 3 HEAA Guidance[431] and our January 2025 Age Assurance and Children's Access Statement (January 2025 Statement),[432] compliance with both the online safety regime and data protection law requirements is mandatory and should not be considered a trade-off. Where providers have concerns or queries around compliance with the data protection regime, they should consult the ICO and their available guidance.[433]

6.58 Where a provider chooses not to implement highly effective age assurance and applies the measures to all users, we do not expect any significant impact on users' rights to privacy. We do not expect providers to collect any further personal data or information about a user to implement the measures using this approach. Any interference with privacy rights should be proportionate as service providers are expected to have regard for protecting users' rights to privacy and comply with data protection laws when they implement safety measures.

6.59 We also determine that where the provider chooses to apply the safety default and support for child users measures to all users it will not significantly impact users' privacy, as we do not expect providers to collect any further personal data on users to undertake this option.

Who this measure applies to

6.60 The amended measures will apply to:

- providers of user-to-user services that have high risk of grooming; and
- providers of large user-to-user services that have medium risk of grooming.

6.61 A number of stakeholders questioned the overall proportionality of the safety default, the supportive information measures, and the options we have recommended for implementation.[434] Some stakeholders said that Option A was disproportionate,[435] with some stakeholders particularly noting the disproportionality for small services, as implementing highly effective age assurance would be complex and costly.[436] Snap, Inc. expressed concerns that blocking all users from network expansion prompts would damage the purpose of social media services and therefore argued Option B was disproportionate for these types of services.[437] One provider submitted that applying these settings to all users would make it harder for users to find people they already know.[438]

6.62 We recognise that measures can impose costs on services. However, we consider that these costs are proportionate given the significant role the measures will play in combatting grooming online. This is especially the case given that we have taken a risk-based approach and targeted the measures at high risk service providers and large medium risk service providers. We consider the risk of a material impact on innovation from these measures is limited given it offers two potential approaches, one of which does not require highly effective age assurance and is therefore likely to have relatively limited up-front costs.

---

[431] See 2.1-2.8 of Part 3 HEAA Guidance.
[432] See 2.23 of January 2025 Statement.
[433] ICO, (n.d.), *UK GDPR guidance and resources*.
[434] [✂]; [✂]; [✂]; [✂]; Ukie response to the June 2025 Consultation, p.38.
[435] British and Irish Law Education Technology Association (BILETA) response to the June 2025 Consultation, p.35; [✂]; Ukie response to the June 2025 Consultation, p.38.
[436] [✂]; Name Withheld 11 response to the June 2025 Consultation, p.2.
[437] Snap Inc. response to the June 2025 Consultation, p.17
[438] [✂].

6.63 We acknowledge the feedback that these measures may involve significant costs if providers were to use highly effective age assurance, or impact user experience on a service if providers applied safety defaults to all users. We proposed two options so that providers can implement the measures in a way that best suits their service to minimise cost or disruption to users.

6.64 We explain in the section 'Effectiveness and benefits' and in the December 2024 Statement the benefits of restricting functionalities that may facilitate grooming.

6.65 Having considered the feedback on who these measures should apply to, our decision remains unchanged from our proposals in the June 2025 Consultation.

Conclusion

6.66 Our analysis shows that the amendments to the Illegal Content Codes discussed in this chapter are likely to materially increase the number of children benefitting from the protections of the measures and increase the number of service providers that are implementing the measures. Given that we expect the measures to be an effective means of combatting grooming, we expect our amendments to result in material reductions in incidents of grooming.

6.67 The analysis above shows that the costs which service providers incur will vary depending on which approach to implementation they take and whether they are already implementing the measures or not. We recognise that in some cases service providers could incur significant costs as a result of this amendment. However, given the severe impact grooming has and the important role these amendments will have in combatting grooming, we have decided that these costs are proportionate. We have therefore decided to proceed with the amendments to the Illegal Content Codes discussed in this chapter.

6.68 The full text of the amended measures can be found in our Illegal Content Codes of Practice for user-to-user services and are referred to as ICU F1 and F2. The measures are part of our Codes of Practice on CSEA duties.

7. Broadening content appeals

What is this chapter about?

This chapter sets out our decision on technical changes that we have made to our Illegal Content and Protection of Children Codes relating to appeals. The changes are designed to extend users' rights of appeal against content moderation decisions. These act as a safeguard on users' freedom of expression.

What decisions have we made?

Our final Code amendments are summarised in the table below:

Number in Codes	Recommended measure	Who should implement this
Various ICU D measures[439]	Broadening the scope of appeals measures to ensure that they cover decisions taken on the basis that content was an 'illegal content proxy'	Various (as set out in the amended measures).
Various PCU D measures[440]	Broadening the scope of appeals measures to ensure that they cover decisions taken on the basis that content was a 'content that is harmful to children proxy'	Various (as set out in the amended measures).

Why have we made these decisions?

These amendments will help ensure users can make content appeals when providers make moderation decisions by identifying proxy content. We consider that the amendments are appropriate and proportionate, as they protect users against excessive moderation and safeguard their rights to freedom of expression while having a limited impact on service providers.

Introduction

7.1 Under section 21(2)(a) of the Online Safety Act 2023 (the Act), user-to-user services have a duty to operate a complaints procedure that allows for "relevant kinds of complaint" to be made.

7.2 In our June 2025 Additional Safety Measures Consultation (June 2025 Consultation), we proposed to make some amendments to the content appeals measures in our Illegal Content and Protection of Children Codes of Practice for user-to-user services (collectively

[439] Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes), Measures ICU D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D12 and D13.
[440] Protection of Children Code of Practice (Protection of Children Code), Measures PCU D2, D3, D4, D5, D6, D7, D8, D9, D10, D13 and D14.

referred to as the Codes)[441] to help safeguard users' human rights.[442] These amendments extended the ability to appeal actions taken by service providers to cover action in relation to "proxies" for illegal content and content harmful to children. In this context, a proxy is "a category of content which is prohibited by a provider's terms of service, which the provider is satisfied includes illegal content or content harmful to children."[443]

7.3 Nothing in the Act requires providers to make illegal content or content harmful to children judgements when moderating content, so long as the application of that provider's own terms and conditions is sufficient to secure compliance with the duties in the Act in other ways. If the provider's own terms and conditions of use prohibit content defined more widely than the definition of illegal content or content harmful to children under the Act, then the provider would be considered to have fulfilled its duties regarding moderation action so long as it applied these terms and conditions properly.

7.4 In this chapter we address the feedback we received on these proposals and put forward our final decisions.

Broadening content appeals to cover proxies

Our approach to the June 2025 proposals

7.5 As noted in our June 2025 Consultation, the Codes contain measures related to "relevant complaints". For example, measures ICU D1 and PCU D1 recommend that providers have systems and processes which enable prospective complainants[444] to make each type of "relevant complaint". The term "relevant complaint" is defined to include the following (emphasis added):

Table 7.1: Definition of 'relevant complaint' in the Codes

Illegal Content Codes for user-to-user services	Protection of Children Code for user-to-user services
Sub-para (c): complaints by a UK 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.	Sub-para (c): complaints by a UK user who has generated, uploaded, or shared content on a service if that content is taken down on the basis that it is content harmful to children.

[441] Illegal Content Codes; Protection of Children Code.
[442] These changes are not needed for search services because the search Codes already have the necessary effect.
[443] This is content that a provider determines to be in breach of its terms of service, in circumstances where these terms are cast broadly enough to necessarily cover illegal content and content harmful to children. In this chapter we refer to content that is an "illegal content proxy" or "content that is harmful to children proxy" as "proxy content".
[444] "Prospective complainants" is defined in each of our user-to-user Codes of Practice as "United Kingdom users" and "affected persons". "Affected person" is defined in section 20 of the Act.

Illegal Content Codes for user-to-user services	Protection of Children Code for user-to-user services
Sub-para (d): complaints by a UK user 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.	Sub-para (d): complaints by a UK user 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 content that is harmful to children.

7.6 The term "relevant complaints" includes a type of complaint which we refer to in the Codes as "content appeals".[445] In our consultation, we noted that "content appeals" did not extend to action that a provider takes on the basis that content was an "illegal content proxy" or a "content that is harmful to children proxy".

7.7 In the Illegal Content Codes for user-to-user services (Illegal Content Codes), we defined "illegal content proxy" as content that a provider determines to be in breach of its terms of service, where: a) the provider has reason to suspect that the content may be illegal content; and b) the provider is satisfied that its terms of service prohibit the type of illegal content which it had reason to suspect existed.[446]

7.8 In our Protection of Children Code for user-to-user services (Protection of Children Code), we define "content that is harmful to children proxy" as "primary priority content proxy, priority content proxy, or non-designated content proxy". This is "content that a provider determines to be in breach of its terms of service, where: a) the provider has reason to suspect that the content may be relevant primary priority content, priority content and/or non-designated content; and b) the provider is satisfied that its terms of service prohibit the type of relevant content which it had reason to suspect existed".[447]

7.9 In this chapter, we refer to content that is an "illegal content proxy" or a "content that is harmful to children proxy" as "proxy content".

7.10 In addition to this change, we also proposed changing the existing defined term "appeal" to "content appeal". This was necessary to distinguish it from an "age assessment appeal" which we explained in chapter 18 of the June 2025 Consultation.

7.11 The limited scope of the content appeals measures meant that where a provider took certain decisions on the basis that a user generated, uploaded, or shared proxy content, the provider would not need to give the user a mechanism to make a content appeal. Those decisions are:

a) a decision to take the content down;
b) a decision to give the user a warning; and

[445] In the Protection of Children Code we refer to these types of complaints as a "content appeal" to distinguish it from an "age assessment appeal". To date, this distinction had not been necessary in the Illegal Content Codes. However, we proposed introducing measures into the Illegal Content Codes about "age assessment appeals". Consequently, we are now making amendments to the Illegal Content Codes to replace all relevant references to "appeals" with "content appeals". We explained this in chapter 18 of the June 2025 Additional Safety Measures Consultation (June 2025 Consultation).
[446] Illegal Content Codes, p.64.
[447] Protection of Children Code, p.29-30 and p.72.

c) a decision to suspend, ban or in any other way restrict the user from using the service.

7.12 We explained that this may have had a negative impact on that user's human rights. For example, in circumstances where a provider decided that a user shared proxy content and took their content down, but the content was not in fact proxy content, the user would not have had a mechanism to bring this to the provider's attention under our existing measures.

7.13 To complement existing and proposed measures and safeguard users' human rights, we proposed in the June 2025 Consultation that users should be able to appeal decisions that providers make on the basis that their content is proxy content.[448] We also considered it important that measures about the action that providers should take after determining a content appeal are broad enough to reflect this position.

7.14 To achieve this objective, we proposed various amendments to the Codes. These amendments fell into two broad categories:

a) Amendments to measures:

i) We proposed to broaden references to "illegal content" to include "or an illegal content proxy", and references to "illegal content judgements" to include "judgements that content is an illegal content proxy" for measure ICU D8.
ii) We proposed to broaden references to "content that is harmful to children" to include "or a content that is harmful to children proxy" for measure PCU D8.
iii) We proposed to broaden the reference to "illegal content" in measure ICU D10 to also include "or an illegal content proxy".
iv) We proposed to broaden the reference to "content that is harmful to children" in measure PCU D10 to include "or a content that is harmful to children proxy".

b) Broadening the scope of content appeals measures to cover content appeals based on proxy content:

i) We proposed to broaden the definitions of "relevant complaint" and "content appeal" so that they apply to decisions made by providers on the basis that content is proxy content.

Legal framework

7.15 Our thinking in the consultation and in this statement is underpinned by a number of provisions in the Act, including our duties to act proportionately and to protect users' freedom of expression and privacy.

[448] In our June 2025 Additional Safety Measures Consultation, we proposed that: (i) users who share, generate or upload child sexual exploitation and abuse (CSEA) content or CSEA content proxies are banned from the service and prevented from returning; and (ii) Providers prepare and apply a sanctions policy in relation to UK users who generate, upload, or share illegal content, illegal content proxy, content harmful to children and/or content that is harmful to children proxy (including primary priority content proxy, priority content proxy and non-designated content proxy.) See chapter 3: User sanctions for more detail.

Providers' duties regarding complaints and safety measures

7.16 Firstly, under section 21(2)(a) of the Act, user-to-user services have a duty to operate a complaints procedure that allows for "relevant kinds of complaint" to be made.[449]

7.17 We consider that where a provider has chosen to moderate by identifying proxy content (rather than making illegal content judgements or determining whether content is harmful to children), it is unlikely to comply with these duties unless it considers content appeals about decisions taken on the basis that content is proxy content.[450]

7.18 Secondly, as explained in paragraphs 21.19 to 21.20 of our June 2025 Consultation, user-to-user services, including those likely to be accessed by children, must take or use proportionate measures to address the risks identified in their illegal content and (where relevant) children's risk assessments. These duties include preventing individuals from encountering priority illegal content, mitigating and managing the risk of priority offences and harm to individuals, and mitigating and managing the risks of harm to children and the impact of such harm to children in different age groups.

7.19 We consider that a provider's decision to moderate by determining proxy content (rather than making illegal content judgments or determining whether content is harmful to children) is unlikely to be proportionate unless it also allows users to make content appeals based on proxy content. This is because of the negative impact on users if they do not have a mechanism to appeal such decisions. For example, it is unlikely to be proportionate to ban a user from a service if they are unable to appeal the decision on which the ban is based. In this context, we note that the Act's objectives are not limited to ensuring that regulated services are safe by design; they also include ensuring that such services are designed and operated in such a way that users' rights are protected.[451]

Ofcom duties regarding measures recommended in the Codes

7.20 As explained in our June 2025 Consultation, we have a duty to prepare and issue one or more Codes of Practice for Part 3 services describing measures recommended for the purpose of compliance with certain duties, including the duties set out in paragraph 7.36.[452] We also have a duty to design measures in the light of certain principles, including the importance of protecting the right of users to freedom of expression.[453] We consider that these amendments ensure that robust safeguards are in place to minimise the impact of our other measures on users' rights to freedom of expression.

---

[449] Sections 21(4)(c) to (d) and 5(c) to (d). Such complaints include: a) 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 or content that is harmful to children; and b) complaints by a user of a service if the provider has: i) given a warning to the user, ii) 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 or content that is harmful to children.

[450] We also note that Category 1 services subject to additional duties regarding complaints which are relevant in this context. These are set out in section 72(6) to (8) of the Act. We also note that they have a duty, when deciding on and implementing safety measures and policies, to have particular regard to the importance of protecting users' right to freedom of expression within the law – see section 22(2) and 10(12) of the Act.

[451] Section 1(3) of the Act.

[452] Section 41(1) to (3) of the Act.

[453] Schedule 4 to the Act, paragraph 10(1) to (2)(a).

Our decision

7.21 We have decided to proceed with the amendments to the measures as proposed in our June 2025 Consultation. The full text of the amended measures can be found in the Codes of Practice.

How this approach works

7.22 We are making these amendments to the content appeals measures in our Codes to help safeguard users' human rights. These amendments will extend the ability to appeal actions taken by service providers to cover action in relation to "proxies" for illegal content and content harmful to children.

7.23 To support this position, we are now confirming the amendments to the measures and broadening the scope of content appeals measures to cover content appeals based on proxy content, as outlined in the following paragraphs.

Amendments to the measures

7.24 We are making several amendments to measures in the Codes. These are set out in Table 7.2.

7.25 Each of the relevant measures applies to a specified class of providers, and we are not changing the class of providers to whom these measures apply.

Table 7.2: Amendments for measures ICU D8.4 and D10.2 and PCU D8.4 and D10.2 to include proxy content[454]

Measure prior to amendment	Summary	Code amendment
ICU D8.4: Appropriate action for relevant complaints which are content appeals – determination (large or multi-risk services).	These measures recommend that providers should have regard to specified matters in determining what priority to give review of content appeals. In ICU D8.4, these matters include (in summary): a) the seriousness of the action taken against the user or in relation to the content 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 content identification technology (and if so, various other matters are specified). c) the past error rate on the service in relation to illegal	ICU D8.4: Broaden references to "illegal content" to include "or an illegal content proxy". Broaden references to "illegal content judgments" to include "judgments that content is an illegal content proxy".
PCU D8.4: Appropriate action for content appeals – determination (services that are large or multi-risk).		PCU D8.4: Broaden references to "content that is harmful to children" to include "or a content that is harmful to children proxy".

[454] For further information on additional amendments made to measure ICU D8, ICU D10, PCU D8, and PCU D10, see chapter 3: User sanctions.

	content judgments of the type concerned. Similar matters are specified in PCU D8.4	
ICU D10.2:[455] Appropriate action for relevant complaints which are content appeals – action following determination.	These measures recommend certain steps that providers should take when they reverse a decision that content was illegal content (in ICU D10.2) or content that is harmful to children (in PCU D10.2).	ICU D10.2: Broaden the reference to "illegal content" to include "or an illegal content proxy".
PCU D10.2: Appropriate action for relevant complaints which are content appeals – action following determination.		PCU D10.2: Broaden the reference to "content that is harmful to children" to include "or a content that is harmful to children proxy."

Amendments to definitions

7.26 We are also introducing measures into the Illegal Content Codes which includes stipulations regarding "age assessment appeals". See chapter 5: Highly effective age assurance in the Illegal Content Codes. Consequently, we are making amendments in the Codes to replace relevant references to "appeals" with "content appeals".

Broadening the scope of content appeals measures to cover content appeals based on proxy content

7.27 We are also broadening the definitions of "relevant complaint" and "content appeal" so that they apply to decisions made by providers on the basis that content is proxy content. As a result, under measures ICU D1 and PCU D1, providers should have systems and processes which enable users to appeal such decisions. Although it will only lead to clarificatory changes in the Codes,[456] these broadened definitions will affect several other measures which relate to content appeals:

a) Measures ICU D2 and PCU D2 recommend that providers have easy to find, easy to access and easy to use systems and processes for relevant complaints (including content appeals).
b) Measures ICU D3 and PCU D3 recommend that certain providers give specified information prior to the submission of a relevant complaint (including content appeals).
c) Measures ICU D4 and PCU D4 recommend that certain providers acknowledge receipt of relevant complaints (including content appeals) and give the complainant an indicative timeframe for deciding the complaint.
d) Measures ICU D5 and PCU D5 recommend that certain providers send further information about how the complaint will be handled upon receipt of relevant complaints.

[455] We have made further changes and are now recommending measure ICU/PCU D10A, for further details see chapter 3: User sanctions.
[456] The changes to this definition only lead to changes to the Codes to differentiate content appeals from age assessment appeals, rather than substantive changes to the appeals process.

e) Measures ICU D6 and PCU D6 recommend that certain providers should enable complainants to opt out from communications following a relevant complaint.
f) Measures ICU D7 and PCU D7 recommend appropriate action for relevant complaints about content which may be illegal content and content harmful to children respectively.
g) Measures ICU D8 and PCU D8 recommend that certain providers should take appropriate action to (among other things):

   i) determine relevant complaints which are content appeals;
   ii) monitor their performance against performance targets relating to the time it takes to determine content appeals and the accuracy of decision making for content appeals; and
   iii) resource itself so as to give effect to those targets.

h) Measures ICU D9 and PCU D9 recommend that certain providers should determine relevant complaints which are content appeals promptly.
i) Measures ICU D10 and PCU D10 recommend steps that providers should take if, in relation to a relevant complaint that is a content appeal, the provider reverses a decision that content was illegal content or content that is harmful to children respectively.
j) Measure ICU D11 which recommends that providers inform complainants of specified matters when they receive relevant complaints which are not content appeals about the use of proactive technology.
k) Measures ICU D12 and PCU D13 which recommend appropriate action for relevant complaints about non-compliance with certain relevant complaints, namely:

   iv) Nominating a responsible individual or team to ensure that such complaints are directed to an appropriate individual or team to be processed; and
   v) Handling relevant complaints in a way that protects UK users (including children in the case of PCU D13) and within timeframes that the provider has determined are appropriate.

l) Measures ICU D13 and PCU D14 (Exception: manifestly unfounded complaints) which recommend that where providers receive relevant complaints that are not content appeals, they may disregard the complaint in certain specified circumstances.

7.28 As a result of our amendments, these measures will apply to content appeals of decisions made by providers on the basis that content is proxy content.

Our reasoning

Effectiveness and benefits

7.29 We consider these amendments are effective in achieving our objectives and will lead to a number of benefits to both providers and users.

7.30 The amendments will ensure that users can make a content appeal when service providers act in a way which may restrict a user's ability to access content and will help to prevent users from being subject to inaccurate content moderation decisions. We consider it is important to protect users' freedom of expression by ensuring that incorrect decisions that content is illegal content proxy and/or content that is harmful to children proxy are corrected by the provider through appropriate action (measures ICU D10 and PCU D10).

7.31 We also identified how a content appeal process is beneficial to service providers, as it can help them to identify and action content that is illegal content and/or content harmful to

children more accurately in the future. We proposed that these benefits should extend to decisions taken on proxy content to protect users against excessive moderation and preserve their rights to freedom of expression. We consider that this may encourage providers to perform a more thorough review in their initial judgements to avoid having to process content appeals.

7.32 In response to our proposals, civil society, industry and a regulatory stakeholder supported our assessment that the amendments as proposed will benefit users, with some noting the protections it affords to users' data protection and right of appeal.[457] One industry stakeholder also identified in its response that it already allows appeals for proxy content in line with the proposals.[458] In light of these responses, we remain of the view that the amendments will be effective at improving user safety online.

7.33 While stakeholder responses have not led us to fundamentally revise our assessment of the benefits of these amendments, several stakeholders provided feedback or requested clarity on specific points of detail. We discuss these points in the following sub-sections.

Legal basis for applying content appeals to proxy content

7.34 In response to our proposals, an industry stakeholder said that we lack a legal basis for proposing the approach. It argued that, under the Act, only Category 1 services need to allow content appeals based on breaches of their terms of service.[459] [460]

7.35 In our June 2025 Consultation, we outlined our legal framework for making these amendments and noted the significant freedom of expression implications that justified the broadening of content appeals to illegal content and content harmful to children proxies. This is summarised in paragraphs 7.45-7.48.

7.36 Category 1 services must allow complaints by users who have generated, uploaded, or shared content if:

a) that content is taken down; or
b) access to it is restricted; or
c) the user has been suspended or banned on the basis that it is "relevant content" (as in, content prohibited by the terms of service).[461]

7.37 In essence, this means that Category 1 services need to allow content appeals of decisions taken on the basis that content was non-compliant with their terms of service, including where their terms prohibit "proxy" content. However, no other services are subject to an equivalent duty. This poses a risk to the freedom of expression rights of users of non-

[457] Children's Commissioners' Office (CCO) response to the June 2025 Consultation, p.24; Digital Resilience in Education response to the June 2025 Consultation, p.16; Evangelical Alliance UK response to the June 2025 Consultation, p.11; ICO response to the June 2025 Consultation, p.22; [✂]; Name Withheld 4 response to the June 2025 Consultation, pp.10-11; National Center for Missing & Exploited Children (NCMEC) response to the June 2025 Consultation p.9; Open Rights Group Campaign response to the June 2025 Consultation, p.43; Pinterest response to the June 2025 Consultation, p.7; [✂ [✂]; We Are Survivors response to the June 2025 Consultation, p.9.

[458] Pinterest response to the June 2025 Consultation, p.7.

[459] Google LLC response to the June 2025 Consultation, pp.36-37.

[460] We consulted on additional duties for Category 1 services, including terms of service, from 10 July 2026 to 2 October 2026.; Ofcom (2026), Consultation: Giving users more choice, control and trust in their online experience.

[461] Section 72(8)(d) of the Act.

Category 1 services, where their content is taken down, or where they have been banned or suspended, based on enforcement decisions in response to proxy content.

7.38 While we accept that (as the industry stakeholder notes) the Act does not extend the same duty to non-Category 1 services, we are not making this recommendation in relation to all terms of service, but only those terms of service which the provider has elected to rely on as illegal content proxies. It is open to any provider which does not wish to consider proxy appeals, to arrange its content moderation process so as to secure that it always makes an illegal content judgement in relation to suspected illegal content. However, if it chooses to use the flexibility we have built into the Codes to make proxy judgements, it should not do so in such a way that removes users' right to appeal in relation to illegal content. The Act must be interpreted in a way which is proportionate, taking into account users' right to freedom of expression. Applying these amended measures to non-Category 1 services is also likely to deliver significant benefits to users of these services by ensuring they have a mechanism to appeal decisions made by providers on the basis that content is proxy content.

7.39 This approach also does not recommend extending content appeals to other additional content (that would not be considered a proxy) which may be moderated for violating a provider's wider terms of service. It only applies to proxy content.

7.40 We consider that the amendments are therefore proportionate for the reasons set out in the following section.

Impacts and costs

7.41 In our June 2025 Consultation, we said that service providers cannot comply with their duties under the Act without a robust content appeals function. We explained they would therefore incur the costs of implementing existing content appeals measures largely because of requirements of the Act. This means that any such costs incurred would not be a result of choices over which we have regulatory discretion.

7.42 We considered that most service providers would have implemented changes to deal with content appeals centrally and would not therefore need to set up additional systems or processes as a result of our amendments. We said that services are likely to use existing systems and processes to deal with these amendments as this was likely to be the most practical and cost-effective approach, meaning that additional costs associated with the amendments were likely to be limited.

7.43 We also said that where a provider has chosen to moderate by identifying proxy content (rather than making illegal content judgements or determining whether content is harmful to children), it was unlikely to comply with its duties unless it considers content appeals about decisions taken on the basis that content is proxy content. Therefore, any additional costs arising would be a direct requirement of the Act.

7.44 While several responses commented on the costs or other impacts of these amendments,[462] these did not provide any evidence that challenged our original costs assessment, which therefore remains unchanged.

[462] See the relevant chapter in annex 1: Further stakeholder responses.

Rights impact

Freedom of expression

7.45 In our June 2025 Consultation, we stated that we considered that the amendments as proposed will safeguard users' right to freedom of expression. We stated that in the absence of the amendments, there is a risk that certain other measures that we recommended may disproportionately interfere with users' right to freedom of expression.

7.46 This may occur where (for example) a provider makes an incorrect decision that content is proxy content and as a result takes action without the user having an ability to appeal the decision. This includes where a provider takes the user's content down, or takes other moderation action (in accordance with content moderation measures)[463] or otherwise bans or sanctions a user.

7.47 We further stated that the amendments we are now confirming will ensure that such users are able to appeal such decisions. We are satisfied that the existence of the content appeal mechanism safeguards users' right to freedom of expression.

7.48 For more detail on how these amendments relate to user sanctions, user banning, and relevant amendments we are confirming alongside those we are confirming, see chapter 3: User sanctions.

Privacy

7.49 In our June 2025 Consultation, we acknowledged that all content appeals systems and processes will involve the processing of personal data of individuals. It will therefore affect users' rights to privacy and their rights under data protection law.

7.50 We assessed the privacy impacts of the measures – prior to these amendments – in our December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement)[464] and our April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement).[465] To the extent there is an additional impact on the privacy of users because of these amendments, we consider this is likely to be proportionate to the benefits that these amendments provide to those users, including the protection they provide for users' rights related to privacy and data protection.

7.51 In the June 2025 Consultation we noted that the duties for a service provider to operate complaints systems and processes that enable relevant complaints are a requirement of the Act, including content appeals. Where a provider has chosen to moderate by identifying proxy content (rather than making illegal content judgements or determining whether content is harmful to children), it is unlikely to comply with these duties unless it considers content appeals about decisions taken on the basis that content is proxy content. On this basis, we consider that these amendments are likely to constitute the minimum degree of interference required to secure that service providers fulfil their illegal content safety duties and child safety duties under the Act.

7.52 We also expect providers to comply with their obligations under data protection law in relation to that processing. This will include implementing measures in a way which

[463] Illegal Content Codes, Measures ICU C1-C10.
[464] Ofcom (2024), December 2024 Statement, Vol.2, chapter Service Design and User Choice, chapter 6: Reporting and complaints.
[465] Ofcom (2025), April 2025 Statement, Vol.4, What should services do to mitigate the risks of online harms to children, p.363-415.

minimises the amount of personal data that is processed (in line with the principle of data minimisation, which our existing measure for manifestly unfounded complaints supports).[466]

Who these amendments apply to

7.53 The amendments relate to existing measures in our Codes, each of which applies to a specified class of providers.

7.54 In response to our June 2025 Consultation, an individual stakeholder and a trade association stakeholder highlighted proportionality concerns with our proposals, arguing that they could place a disproportionate burden on smaller services.[467]

7.55 We consider that these amendments are proportionate for service providers of all sizes. As set out in paragraph 7.42, providers are likely to be able to use existing processes to deal with additional content appeals resulting from these amendments, therefore incurring limited additional costs. Where such costs are incurred, these can be considered a requirement of the Act.

7.56 Therefore, our assessment about who these amended measures should apply to is unchanged and we are recommending that the types of providers that these amended measures apply to remains the same as in our Codes.

Conclusion

7.57 Our analysis shows that these amendments will provide benefits to users by safeguarding their right to freedom of expression, while having a limited or negligible impact on service providers. Overall, we conclude the amendments are proportionate in that the benefit they will provide significantly outweighs any likely impact on providers.

7.58 We have therefore decided to proceed with the amendments to the measures as proposed in our June 2025 Consultation for the reasons outlined in this chapter. The full text of the amended measures can be found in our updated Illegal Content Codes and our Protection of Children Code.

[466] Illegal Content Codes, Measures ICU D13.
[467] Association for UK Interactive Entertainment (Ukie) response to the June 2025 Consultation, pp.39-40; [✂].

8. Introduction to Proactive Technology

What is this chapter about?

In this chapter we explain what proactive technology is and provide more detail about our constraints when making recommendations about proactive technology under the Act. We also respond to stakeholder feedback related to our overarching approach for proactive technology measures.

Introduction to this section

8.1 The regulatory framework set up by the Online Safety Act 2023 (the Act) is designed to make life online safer for UK users by reducing the risks posed by illegal content and content harmful to children. Proactive technology is an important tool that service providers can use to assist in detecting this content. Without proactive technology, providers are solely reliant on methods like human moderation and user reporting. Given the volumes of content on services, these methods alone are unlikely to be sufficient to protect UK users.

8.2 In our June 2025 Additional Safety Measures Consultation (June 2025 Consultation), we proposed five new proactive technology measures. Two of the proposed measures recommended proactive technologies assessed by the provider. A further three measures recommended the use of a specific type of proactive technology called 'hash matching'. We confirmed one of the hash matching measures in our May 2026 Detecting Intimate Image Abuse Statement (May 2026 Statement) and the others in this statement (see chapter 10: Hash matching).[468]

8.3 This chapter expands on what we said in the June 2025 Consultation, addressing what proactive technology is and providing more detail about our constraints when making recommendations about proactive technology under the Act. We also respond to stakeholder feedback related to our overarching approach for the proactive technology measures before summarising our decisions.

What is proactive technology?

8.4 We use the term 'proactive technology' to describe a set of technologies that can be used as part of a moderation system. These technologies, as defined in the Act, are as follows:

- **Content identification technology:** 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).[469]

[468] Ofcom (2026), May 2026 Statement, Detecting Intimate Image Abuse Statement, p.7-58.
[469] Content identification technology is not regarded as proactive technology if it is used in response to a report from a user or other person about particular content. (Section 231(2) and (3)).

- **User profiling technology:** technology which analyses any, or all of, relevant content, user data, or metadata relating to relevant content or user data for the purposes of building a profile of a user to assess characteristics such as age.[470]
- **Behaviour identification technology:** technology which analyses any, or all of, relevant content, user data, or metadata relating to relevant content or user data to assess a user's online behaviour or patterns of online behaviour (for example, to assess whether a user may be involved in, or a victim of, illegal activity).[471]

8.5 Within these sub-sets of proactive technology as defined by the Act there are a range of different types of technology including rules-based software such as hash matching or text matching, automated content classifiers (algorithmic systems which analyse content and associated signals), and network analysis.[472]

8.6 We proposed three measures in the June 2025 Consultation which recommended the use of the specific type of proactive technology called 'hash matching'.

What is hash matching?

8.7 Hash matching refers to a process for detecting content that has previously been identified as being illegal, or otherwise violative, to prevent the upload and sharing of that content. It involves creating a hash (identifying a series of characters) of a unique piece of content, storing that hash in a database, and using an algorithm to detect attempts to upload the same or similar versions of that content. Once a match of some form is established, the content can either undergo review – which may include human review – or be removed automatically.[473]

8.8 There are two basic forms of hash matching:

- **Cryptographic hash matching** – only detects files that are an exact match. This means that perpetrators may slightly alter images to circumvent it.
- **Perceptual hash matching** – determines whether a given file is perceptually similar to a hashed file by setting a threshold that is used to determine when there is sufficient similarity between the hashes to be considered a (near) match. This allows modifications from an original file to be detected.

8.9 Cryptographic and perceptual hashes are generated using different types of hashing algorithms and are not interchangeable. As such, cryptographic hashes cannot be used for similarity-based matching such as perceptual hash matching.

---

[470] Technology which a) analyses data specifically provided by a user for the purposes of a provider verifying or estimating the user's age in order to decide whether to allow the user to access a service (or part of a service) or particular content; and b) does not analyse any other data or content, is not regarded as a user profiling technology. Section 231(4) and (5).

[471] This technology is not regarded as proactive technology if it is used in response to concerns identified by another person or an automated tool about a particular user. (Section 231(6) and (7)).

[472] We refer to the range of inputs that proactive technology may analyse (such as relevant content and, in the case of user profiling and behaviour identification technology, user data and metadata relating to relevant content or user data) as 'inputs'.

[473] The requirements for human review are set out in annex 4.3: Human review for proactive technology.

8.10 The types of content that can be hashed include images, videos or files that are generated, uploaded to, or shared on, the providers' services. This can include content generated by artificial intelligence.

Analytical framework for making recommendations regarding the use of proactive technology

8.11 Proactive technology is capable of quickly analysing a large volume of content. It is therefore important to ensure that such technology is used by service providers responsibly and in a way which detects target illegal content and content harmful to children but does not disproportionately impact the rights of users.

8.12 There are several factors that the Act requires us to consider when recommending any type of proactive technology. These are reviewed in turn in the following sub-sections.

Content communicated privately

8.13 The Act states that Ofcom may not recommend in a Code of Practice the use of proactive technology to analyse user-generated content communicated 'privately', or metadata relating to that content.[474] For this reason, the measures set out in the following chapters only apply to content communicated publicly by means of a provider's service. We address stakeholder feedback related to this in the relevant section in annex 1: Further stakeholder responses.

Technical feasibility

8.14 The Act requires Ofcom to have regard to the principle that the measures described in a Code of Practice must be technically feasible.[475] We do not consider that it would be technically infeasible to implement proactive technology merely because to do so would require some changes to be made to the design of the service, its operation or both.[476]

8.15 Where a service provider claims that it is technically infeasible for it to implement proactive technology for our measures, and we have concerns about this, we have the power to investigate. Should we find that it is technically feasible for the provider to implement the proactive technology, the relevant measure will apply to them. We do not consider that technical limitations will necessarily remain on an ongoing basis. We address stakeholder feedback related to this in the relevant section in annex 1: Further stakeholder responses.

Accuracy, effectiveness and lack of bias

8.16 The Act requires Ofcom to have regard to the degree of accuracy, effectiveness and lack of bias achieved by the proactive technology in question in deciding whether to recommend it in a Code of Practice or use it as part of a confirmation decision.[477] 'Accuracy',

[474] Schedule 4 to the Act, paragraphs 13(4). We have set out guidance which is intended to assist services in deciding whether content is communicated 'publicly' or 'privately' for this purpose: Guidance on content communicated 'publicly' and 'privately' under the Online Safety Act.
[475] Schedule 4 to the Act, paragraph 2(c).
[476] We explain this in further detail in the section entitled 'End-to-end encryption' in the relevant section in annex 1: Further stakeholder responses.
[477] Schedule 4 to the Act, paragraph 13(6).

'effectiveness' and 'lack of bias' are not defined in the Act. We take each to have their ordinary meanings in English:

a) 'Accuracy' is to do with 'being exact or correct'.
b) 'Effectiveness' is to do with being 'successful in producing a desired or intended result'.
c) 'Lack of bias' is to do with being equitable and fair.

8.17 We note that the concepts are inherently linked. For example, a technology which is biased or inaccurate will be ineffective.

8.18 In our proposed Draft Proactive Technology Guidance as part of the June 2025 Consultation, we took these concepts and refined them further to explain how a regulated service provider could assess them in considering the deployment of technology recommended in Codes of Practice:

- Accuracy: the technical correctness and reliability of proactive technology as assessed during development, testing, and evaluation, including the extent to which errors are minimised through evidence-based methodologies and appropriate performance metrics.
- Effectiveness: the utility and ability of proactive technology to achieve intended outcomes, evaluated by assessing its real-world impact, alignment with specific goals, and operational relevance for the intended use case.
- Lack of bias: the extent to which proactive technology avoids discrimination and negative impacts on equity and fairness, including with respect to the treatment of different users and the handling of different types of content or other inputs.

8.19 In annex 4.1: Proactive technology: CSAM proactive technology technical annex, we describe in more detail how we assessed proactive technology against these concepts for the purposes of recommending specified technologies as part of a measure in a Code of Practice:

- In considering accuracy, we evaluated the available evidence relating to the extent to which the core capabilities of approaches within each technology class support correct identification of the target harm, including the relevance of the signals we expect to be available to the technology and the potential for detection errors.
- In considering effectiveness, we evaluated the available evidence relating to the extent to which the core capabilities of approaches within each technology class support the intended detection objective, taking account of their strengths, limitations and known constraints.
- In considering lack of bias, we evaluated the available evidence relating to the extent to which, based on their core capabilities, approaches within each technology class appeared capable of operating without creating systematic differences in outcomes across relevant groups, languages, contexts or content types.

8.20 We set out our evidence for the accuracy, effectiveness and lack of bias of the technology we are recommending in the following chapters:

- For our measure on proactive technology to detect image-based CSAM, see annex 4.1: Proactive technology: CSAM proactive technology technical annex.
- For the measure on perceptual hash matching for terrorism content, this can be found in the section entitled 'Measures recommending the use of hash matching' in chapter 10: Hash matching.

- For the measure on hash matching to detect and remove known image-based CSAM, see the section entitled 'Automated content moderation' in our December 2024 Statement on Protecting People from Illegal Harms (December 2024 Statement).

Other requirements

8.21 Ofcom has a duty to carry out its functions consistently with human rights. Many human rights may be relevant, including the rights of victims and the right to property. However, the two main rights engaged by proactive technology are freedom of expression and privacy. In this section, we outline the relevant human rights requirements, before considering feedback received in response to the June 2025 Consultation that is relevant to all our proactive technology measures, and setting out our response.

Human Rights

Freedom of expression

8.22 Article 10 of the European Convention of Human Rights (ECHR) requires Ofcom to ensure its measures are compatible with users' and service providers' freedom of expression, such that any interference with freedom of expression must be prescribed by law and necessary and proportionate for a legitimate aim.

Privacy

8.23 Article 8 of the ECHR requires Ofcom to ensure its measures are compatible with individuals' right to privacy, such that any interference with a person's privacy must be in accordance with the law and necessary and proportionate for a legitimate aim. For further considerations relating to freedom of expression and privacy impacts, see the sections entitled 'Rights Impact' in chapter 9: Deployment of proactive technology to detect image-based CSAM and chapter 10: Hash matching.

Stakeholder responses to our June 2025 Consultation

8.24 We received consultation responses from a variety of stakeholders about the human rights implications of the measures, including their impact on freedom of expression and privacy, as well as concerns that the measures could amount to general monitoring. While some responses focused on individual measures, the issues raised are relevant across our measures and are therefore addressed collectively here. We set out the feedback in more detail below and our response in the next sub-section.

8.25 Respondents to the Open Rights Group Campaign expressed concerns that, even if the technology worked reliably, it could infringe upon user rights under UK GDPR.[478]

8.26 Two civil society stakeholders, a campaign affiliated with a civil society stakeholder, and one industry stakeholder, argued that our measures could amount to general monitoring.[479] Google LLC argued that the extension of harms, for which the measures require proactive technology to scan, is disproportionate and risks introducing general monitoring obligations

[478] Respondents to the Open Rights Group Campaign response to the June 2025 Additional Safety Measures Consultation, p. 2.

[479] European Digital Rights Organisation response to the June 2025 Additional Safety Measures Consultation, p. 1; Google LLC response to the June 2025 Additional Safety Measures Consultation, pp. 4, 13-14; Open Rights Group response to the June 2025 Additional Safety Measures Consultation, pp. 15-16, 20; Respondents to the Open Rights Group Campaign response to the June 2025 Consultation, p. 2.

to service providers in a way that Parliament could not have intended.[480] Open Rights Group, respondents to the Open Rights Group Campaign and Google LLC also argued this could disproportionately impact users' privacy and freedom of expression.[481]

8.27 Several stakeholders sought clarity on how measures recommending the use of proactive technology coincide with Article 8 of the European Union's (EU) Digital Services Act (DSA). Article 8 of the DSA states that the DSA imposes no general monitoring or general proactive fact-finding obligations in respect of illegal content on service providers. Three civil society stakeholders, including a campaign affiliated with a civil society stakeholder, argued that the proactive technology measures are disproportionate and could make it difficult for providers to comply with EU law or potentially lead to providers exiting the UK market or both.[482] Open Rights Group warned Ofcom's growth duty could conflict with recommending measures that impose general monitoring, creating legal and trade risks.[483] European Digital Rights Organisation added that the Act would require EU service providers to possibly implement indiscriminate content monitoring, which it stated would conflict with EU law prohibiting general monitoring.[484]

Our response to stakeholder feedback on human rights and general monitoring

8.28 Ofcom recognises the points that stakeholders raised about general monitoring, including the impact on user privacy and proportionality, which we discuss further below. However, we place little weight on the arguments from civil society stakeholders that these measures make it harder for service providers to comply with EU laws such as the DSA. We received no stakeholder feedback from service providers making this point. In general, there is already widespread use of proactive technology as part of content moderation practices, and it is not clear to us that our recommendations would make it materially harder to comply with EU law. The measures described in a Code of Practice relate only to the design or operation of a Part 3 service in the UK, or as it affects UK users of the service, and as such it may be possible for providers to use proactive technology in line with our measures to analyse content accessible to UK users, but not for users in other jurisdictions where different laws may apply.[485] [486] [487]

8.29 While some stakeholder comments argued that content identification technology is inherently inconsistent with user privacy, we note that Parliament has expressly conferred on Ofcom a power to recommend proactive technology in the Act where this is proportionate to comply with the illegal content or children's safety duties (the safety duties) and is not limited to particular kinds of content. The Act sets out the conditions under which Ofcom may recommend such proactive technology in Schedule 4.[488] The Act

---

[480] Google LLC response to the June 2025 Consultation, pp. 4, 13-14.

[481] Google LLC response to the June 2025 Consultation, pp. 4, 13-14; Open Rights Group response to the June 2025 Consultation, pp. 15-16, 20; Respondents to the Open Rights Group Campaign response to the June 2025 Consultation, p. 2.

[482] European Digital Rights Organisation response to the June 2025 Consultation, p. 1; Open Rights Group response to the June 2025 Consultation, pp. 15-16, 20; Respondents to the Open Rights Group Campaign response to the June 2025 Consultation, p. 2.

[483] Open Rights Group response to the June 2025 Consultation, pp. 15-16, 20.

[484] European Digital Rights Organisation response to the June 2025 Consultation, p. 1.

[485] Schedule 4 to the Act, paragraph 11.

[486] Schedule 4 to the Act, paragraph 13(3).

[487] Schedule 4 to the Act, paragraph 10(1) & 10(2).

[488] Schedule 4 to the Act, paragraph 13(3).

requires Ofcom, when recommending measures in a Code of Practice, to design those measures in light of the importance of protecting the privacy of users and incorporating privacy safeguards.[489] Parliament has therefore explicitly contemplated Ofcom recommending this kind of technology and did not limit Ofcom's powers in the way that stakeholders suggest. As such, we do not consider content identification technology, in itself, amounts to general monitoring that infringes user privacy. For these reasons, we consider that UK law permits the use of proactive technology when it is proportionate and in accordance with the restrictions set out on Ofcom's powers in line with the Act.[490]

The distinction between proactive technology and accredited technology

8.30 Ofcom's powers in relation to use of accredited technology to address CSEA content or terrorism content under section 121 of the Act are distinct from its powers to recommend measures for the use of proactive technology in a Code of Practice. This chapter and the next (and the Draft Proactive Technology Guidance) relate to the use of proactive technology, not the use of accredited technology. Ofcom's published materials about proactive technology and accredited technology remain separate, as the two frameworks differ.

8.31 Under paragraph 13 of Schedule 4 to the Act, Ofcom may recommend the use of proactive technology in a Code of Practice where it is a proportionate and technically feasible way of complying with the safety duties, where Ofcom has had regard to the accuracy, effectiveness and lack of bias of the technology in question. Such a recommendation only applies to content communicated publicly and does not create a mandatory obligation on providers to use that proactive technology. On the other hand, and in relation to section 121 of the Act, a specific technological product may be accredited as meeting the minimum standards of accuracy (as set by the Secretary of State) and therefore capable of being required for use by a provider in a Technology Notice issued by Ofcom.[491] The purpose of the accredited technology framework is only intended to determine whether a technology could be considered for requirement by Ofcom through a Technology Notice to address only terrorism content communicated publicly, or CSEA content communicated publicly or privately. A technology could potentially satisfy one framework without necessarily satisfying the other, because the frameworks are directed at different statutory objectives, apply in different legal contexts, and assess different forms of technology for different deployment contexts.

Overview of what we proposed

8.32 In our June 2025 Consultation, we consulted on various proactive technology measures, including:

a) Measures proposing proactive technologies assessed by the provider:

[489] Schedule 4 to the Act, paragraph 10(1) & 10(2).
[490] We explain in further detail why this is proportionate in the section entitled 'Impact and costs' in chapter 9: Deployment of proactive technology to detect image-based CSAM.
[491] Department for Digital, Culture, Media & Sport, 2026. Minimum standards of accuracy in the detection of terrorism and CSEA content.

i) Measures for providers to assess and deploy proactive technology to support the detection of target illegal content (ICU C11/C12);[492]
ii) Measures for providers to assess and deploy proactive technology to support the detection of content harmful to children (PCU C9/C10).[493]

b) Hash matching measures:

i) A measure recommending the use of perceptual hash matching for terrorism content (ICU C13);
ii) An amendment to our perceptual hash matching measure for known image-based CSAM (child sexual abuse material) to extend the number of high-risk services in scope (ICU C9); and
iii) A measure recommending the use of perceptual hash matching for intimate image abuse (ICU C14), which we have since confirmed in our May 2026 Statement.

8.33 With regard to our measures proposing proactive technologies assessed by the provider, we are now recommending one measure rather than two and confirming this amended version for unknown image-based CSAM only (ICU PT1). We define unknown image-based CSAM as content that is not included in any reference database of hash values for CSAM used by the provider in order to comply with ICU C9 of the Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes). We acknowledge that proactive technology that is designed and deployed to detect unknown CSAM is also capable of detecting known CSAM. However, to ensure the detection and removal of CSAM is as quick and effective as possible, we consider it is proportionate that in-scope services continue to use our recommended hash matching for known CSAM (ICU C9) alongside proactive technology for unknown CSAM.

8.34 We aim to consult on an amended version of Measure ICU PT1 for additional harms in Spring 2027. We explain our reasoning for this in chapter 9: Deployment of proactive technology to detect image-based CSAM.

8.35 We have decided to proceed with our hash matching measures. We explain our reasoning for this in chapter 10: Hash matching.

8.36 In response to concerns raised by stakeholders around the risk of false positives, we are also recommending measures on steps providers should take to reverse incorrect content decisions. We set out more detail on this in chapter 9: Deployment of proactive technology to detect image-based CSAM – Steps to take on reversal of content decisions.

[492] Target illegal content means relevant content that amounts to an offence in relation to the relevant harms; or is illegal content proxy, where the provider is satisfied that its terms of service prohibit the relevant harms.
[493] Target content harmful to children means relevant content that either: a) relates to one of the following: i) where the service is a large service, the specific kinds of primary priority content for which the service is at medium or high risk, to the extent they are relevant primary priority content; or ii) where the service is not a large service, the specific kinds of primary priority content for which the service is at high risk, to the extent they are relevant primary priority content; or b) is content that is harmful to children, where the provider is satisfied that its terms of service prohibit the specific kinds of relevant primary priority content.

9. Deployment of proactive technology to detect image-based CSAM

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

What is this chapter about?

This chapter discusses the use of proactive technology[494] to detect previously unknown image-based child sexual abuse materials (CSAM).

Proactive technology guidance

Alongside this measure we are also publishing our draft proactive technology guidance to give service providers direction on how to implement proactive technologies.

What decisions have we made?

We have decided to introduce a measure into our Codes of Practice[495] recommending service providers use recommended proactive technology for the purpose of detecting unknown image-based CSAM. We have assessed the degree of accuracy, effectiveness and lack of bias achieved by those technologies if configured and deployed to meet the proactive technology criteria we recommend.

We intend to consult on proactive technology for further harms in Spring 2027.

Our final Code measures are summarised in the table below:

Number in Codes	Recommended measure	Who should implement this
ICU PT1	Deployment of proactive technology to detect image-based CSAM content.	Providers of: a) Large user-to-user services at medium risk of image-based CSAM. b) User-to-user services with more than 700,000 monthly active users in the UK at high risk of image-based CSAM. c) File storage and file sharing services at high risk of image-based CSAM.
ICU C4a	Steps to take on reversal of content decisions.	All user-to-user services.

[494] Proactive technology describes technology that can be used as part of a content moderation system. The technology analyses content, user data and metadata to identify a range of outcomes including types of content, for example illegal content, user profiles, for example identifying user age, user behaviour, for example if a user is involved in illegal activity.

[495] Illegal Content Code of Practice for user-to-user services: Proactive Technologies (unknown content)

Number in Codes	Recommended measure	Who should implement this
PCU C4a	Steps to take on reversal of content decisions.	All user-to-user services likely to be access by children.
ICS C3A	Steps to take on reversal of content decisions.	All search services.
PCS C4A	Steps to take on reversal of content decisions.	All search services likely to be accessed by children.

Why have we made these decisions?

Image-based CSAM can have a profound and long-lasting impact on children who are sexually abused, and on the wellbeing of individuals who encounter this content. The scale of online CSAM is increasing year on year,[496] and urgent action is needed to tackle this.

Our analysis shows that, when the recommended proactive technology meets the proactive technology criteria, it is capable in principle of being accurate, effective and lacking in bias in detecting image-based CSAM. The use of proactive technology should significantly increase the amount of CSAM that is detected and removed. This will help reduce the re-traumatisation of victims/survivors of abuse, improve law enforcement agencies' ability to disrupt ongoing abuse and reduce the risk of perpetrators abusing more children in future.

In response to stakeholder feedback, we have made the proactive technology measure more specific since the consultation and carried out a more detailed assessment of the accuracy and effectiveness of the technology.

Introduction

9.1 In this chapter, we set out our decision to include a measure in our Draft Illegal Content Code of Practice for User-to-User Services: Proactive Technologies (unknown content)[497] specifying that providers should use proactive technology for the purpose of detecting unknown image-based child sexual abuse material (CSAM) as it manifests on their service.[498] In doing so, providers should use one or a combination of the technologies set out in our measure. They should ensure that in selecting these technologies, and throughout their lifecycle, they meet the proactive technology criteria. We have assessed that the technologies set out in our measure are capable in principle of being accurate, effective and lacking in bias when providers do this.

9.2 The structure of this chapter is as follows:

i) The first section sets out a summary of what we proposed in the June 2025 Additional Safety Measures Consultation (June 2025 Consultation), the feedback we received, and our decision based on this feedback.

[496] Internet Watch Foundation, 2025. Executive Summary, IWF Annual Data & Insights Report 2025.
[497] We will refer to this as the Illegal Content Code: Proactive technology throughout this chapter.
[498] Image-based CSAM is Content amounting to an offence specified in any of the following paragraphs of Schedule 6 to the Act: 2 to 4, 7 or 10, so far as the risk in relation to those offences relates to CSAM in the form of photographs, videos or visual images. Unknown image-based CSAM is Image-based CSAM content that is not included in any reference database of hash values for CSAM used by the provider in order to comply with ICU C9 of the Illegal content Codes of Practice services for user-to-user services (Illegal Content Codes).

ii) The second section explains in more detail how this measure works.
iii) We then go on to outline our reasoning in more detail, including the benefits and effectiveness of the measure, benefits and effectiveness of the proactive technology criteria and guidance, costs, rights impacts and our reasoning for who the measure applies to, before outlining our overall conclusion.

What we proposed in the consultation

9.3 In our June 2025 Consultation, we proposed criteria for providers to assess both existing and new proactive technology for detecting target illegal content[499] and content harmful to children.[500] We proposed that service providers should assess whether proactive technology that met these criteria existed. We proposed that where providers assessed that proactive technology met all these criteria, it should be deployed to detect relevant harms on their service.

9.4 The criteria we proposed providers should assess proactive technology against are outlined in the following table.

Table 9.1: Proactive technology criteria from June 2025 Consultation

Criteria	Description
Use of high-quality data	The proactive technology has been developed and tested using high-quality[501] datasets appropriate to and reflecting a broad range of inputs relevant to the harm[502] it is intended to detect (as identified in the service's risk assessment(s)).
Addressing biases	Potential biases have been identified and addressed during the design and development process, and risks are appropriately managed and addressed throughout the proactive technology's lifecycle.

[499] We proposed target illegal content should include the following harms: Image-based CSAM; CSAM URLs; grooming; fraud (and financial services); encouraging or assisting suicide (or attempted suicide).
[500] We proposed target content harmful to children should include the following harms: primary priority content.
[501] "High-quality" refers to data that is accurate, legally and ethically sourced, well-labelled (where appropriate), representative of the harm and context the proactive technology is intended to address and sufficiently diverse to support meaningful evaluation, including (where relevant) testing the proactive technology's ability to detect content it has not previously encountered. The degree to which a dataset may be considered "high-quality" will vary based on the harm being addressed and the context in which the proactive technology is used. However, indicators that a dataset might not be considered "high-quality" would include (for example) unclear sourcing, evidence of poor or inconsistent labelling, or limited representation of relevant harms or input types.
[502] We proposed this could be a relevant proxy, if applicable.

Criteria	Description
Evaluating performance	The proactive technology has been evaluated using appropriate performance metrics and configured so that its performance strikes an appropriate balance between precision[503] and recall.[504] [505]
Safeguards against misuse and exploitation	Safeguards are in place to identify and appropriately manage security threats and risks of exploitation and misuse, including through the use of access restrictions[506] and system integrity protections.
Contextual testing and evaluation	The proactive technology's performance has been evaluated in real-world use cases relevant to the provider's content (having regard to the risk of harm to individuals identified in the providers' risk assessment(s)) and the results indicate it correctly detects the harm it is intended to detect. This includes testing for scalability, handling of different media types (where relevant), and whether the proactive technology's performance could be improved by layering with complementary approaches or (in the case of existing deployments) by updating to a more current version.
Maintenance and ongoing monitoring	Mechanisms are in place to monitor and maintain the proactive technology's effectiveness over time, including processes for regular review and iterative adjustments to respond to emerging circumvention techniques, biases, or new content types.
Human review	Policies and processes are in place for human review and action is taken in accordance with that policy, including the evaluation of outputs during development (where applicable), and the human review of an appropriate proportion of the outputs of the proactive technology during deployment. Outputs should be explainable to the extent necessary to support meaningful human judgement and accountability.[507]

[503] We explained in our June 2025 Consultation that precision is the proportion of identified cases that are true positives.

[504] We explained in our June 2025 Consultation that recall is the proportion of true positive cases that are correctly identified.

[505] Please see the section entitled 'Additional factors to consider for 'Evaluating performance'' and 'Human review' for certain factors providers should consider related to the testing, configuration, deployment and ongoing monitoring of the proactive technology.

[506] "Access restrictions" refers to limitations placed on interaction with system operations or data, based on the roles or responsibilities assigned to any entity (including individuals, devices, or systems such as software components, applications, and automated processes), whether internal and external. This aligns with principles on controlling access, as outlined in the UK Cyber Assessment Framework, and supported by international standards such as ISO/IEC27001 and NIST SP-800-53.

[507] Please see the section entitled 'Additional factors to consider for 'Evaluating performance' and 'Human review' for certain factors providers should consider related to the testing, configuration, deployment and ongoing monitoring of the proactive technology.

Criteria	Description
Incorporating feedback	Feedback mechanisms are in place to maintain or improve performance over time. This includes updating proactive technology with diverse and up-to-date datasets to reflect evolving trends or emerging types of illegal content and/or content harmful to children and/or integrating ongoing feedback from users and individuals working in content moderation[508] into its development, while managing the risk of introducing additional bias.

9.5 We also provided additional factors for providers to consider in relation to the criterion on evaluating performance and human review.[509]

9.6 We proposed 'Proactive Technology Draft Guidance' which included further guidance on the proactive technology criteria, including examples of how providers could assess proactive technology against the criteria.[510]

9.7 We proposed that these measures would apply to providers of the following services:

- Large user-to-user services that are medium/high risk for at least one relevant harm;[511]
- User-to-user services with more than 700,000 monthly UK users that are high risk for at least one relevant harm;[512]
- User-to-user services that are file-storage and file-sharing services which identify a high risk of image-based CSAM regardless of size; and
- All user-to-user services which identify a high risk of grooming.

Summary of stakeholder feedback

9.8 Several stakeholders, including those from civil society, law enforcement, industry and academia expressed their support for the proposals.[513] Some stated that the proposals

[508] For an explanation of 'individuals working in content moderation', see our December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement): Volume 2: Service design and user choice, pp. 79-80.
[509] One stakeholder provided support for the additional factors. Information Commission's Office (ICO) response to the June 2025 Additional Safety Measures Consultation, p. 10.
[510] One stakeholder expressed support for our Draft Proactive Technology Guidance. OnlyFans response to the June 2025 Additional Safety Measures Consultation, (email correspondence).
[511] Where this harm is a relevant harm harmful to children, this applies to such services which are likely to be accessed by children.
[512] Where this harm is a relevant harm harmful to children, this applies to such services which are likely to be accessed by children.
[513] Barnardo's response to the June 2025 Additional Safety Measures Consultation, p. 6, 8; British and Irish Law Education and Technology Association (BILETA) response to the June 2025 Additional Safety Measures Consultation, p. 7; Canadian Centre for Child Protection response to the June 2025 Additional Safety Measures Consultation, p. 9; Check My Ads Institute response to the June 2025 Additional Safety Measures Consultation, p. 4; Children's Commissioner's Office (CCO) response to the June 2025 Additional Safety Measures Consultation, p. 10; Common Sense Media response to the June 2025 Additional Safety Measures Consultation, p. 5; Friends2Follow (F2F) response to the June 2025 Additional Safety Measures Consultation, p. 2; Internet Matters response to the June 2025 Additional Safety Measures Consultation, p. 8; National Centre for Missing & Exploited Children (NCMEC) response to the June 2025 Additional Safety Measures Consultation, p. 2; National Police Chief's Council (NPCC) response to the June 2025 Additional Safety Measures

could improve the identification of relevant content before harm occurs and that they reflected current industry practice.[514] Some stakeholders expressed their support specifically in relation to protecting children from being exposed to harmful content online.[515]

9.9 Several stakeholders also expressed their support for the approach taken to the measures, including their 'principles-based' nature, their ability to keep pace with technological evolution and the use of criteria to mitigate negative rights impacts.[516] Some providers encouraged further flexibility in both the measures and associated guidance and criteria.[517] [518]

9.10 However, several stakeholders criticised our measures and said that they could have a negative impact on users' rights. They raised the following points:

- Our measures were too flexible, leading to the outcome of the measures not being sufficiently foreseeable for users as it varies from service to service. Some of these stakeholders argued that this meant the measure was incompatible with the European Convention on Human Rights (ECHR).[519]

Consultation, pp. 3-4; Online Dating and Discovery Association's response to the June 2025 Additional Safety Measures Consultation, p. 3; Which? response to the June 2025 Additional Safety Measures Consultation, p. 1

[514] Internet Matters response to the June 2025 Consultation, p. 8; Online Dating and Discovery Association's response to the June 2025 Consultation, p. 3.

[515] CCO response to the June 2025 Consultation, p. 10; Digital Resilience in Education - Welsh Government response to the June 2025 Additional Safety Measures Consultation, pp. 7-8; Internet Matters response to the June 2025 Consultation, p. 8; [✂].

[516] BILETA response to the June 2025 Consultation, pp. 7-8, 10, 12; Glitch response to the June 2025 Additional Safety Measures Consultation, p. 2; [✂]; ICO response to the June 2025 Consultation, p. 9; Molly Rose Foundation response to the June 2025 Additional Safety Measures Consultation, p. 6, 13; Online Safety Act Network response to the June 2025 Additional Safety Measures Consultation, p. 11; [✂]; Refuge response to the June 2025 Additional Safety Measures Consultation, p. 5.

[517] Computer and Communications Industry Association response to the June 2025 Additional Safety Measures Consultation, p. 6; Google LLC response to the June 2025 Additional Safety Measures Consultation, p. 17; [✂]; [✂]; [✂]; [✂]; [✂]; [✂]; [✂].

[518] Google LLC response to the June 2025 Consultation, p. 17; Meta Platforms Inc. (Meta) response to the June 2025 Additional Safety Measures Consultation, p. 10.

[519] A member of the Philippine Survivor Network response to the June 2025 Additional Safety Measures Consultation, pp. 11-12; BILETA response to the June 2025 Consultation, pp. 8-9;; Gender and Tech Research Lab's response to the June 2025 Additional Safety Measures Consultation, p. 4; Internet Watch Foundation response to the June 2025 Additional Safety Measures Consultation, p. 18, 23; Microsoft response to the June 2025 Additional Safety Measures Consultation, p. 3; [✂]; NPCC response to the June 2025 Consultation, p. 4; National Society for the Prevention of Cruelty to Children (NSPCC) response to the June 2025 Additional Safety Measures Consultation, pp. 2-4, 26; Open Rights Group response to the June 2025 Additional Safety Measures Consultation, pp. 16-17, 19-20, 25; Professor Korff, D. response to the June 2025 Additional Safety Measures Consultation, p. 4, 6; Respondents to the Open Rights Group Campaign response to the June 2025 Additional Safety Measures Consultation, pp. 2, 5-6; Smith, G. response to the June 2025 Additional Safety Measures Consultation, pp. 1-9; South West Grid for Learning response to the June 2025 Additional Safety Measures Consultation, p. 7; [✂]; [✂]; The Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Additional Safety Measures Consultation, p. 5; UK Finance response to the June 2025 Additional Safety Measures Consultation, p. 9; [✂].

- Our measures may amount to prior restraint of free expression, particularly where fully automated systems suppress content without human review.[520] Some of these stakeholders also argued that harmful behaviour should be identified before content is uploaded or shared,[521] while others highlighted that the effectiveness of proactive technology may vary across languages and dialects, potentially having a discriminatory impact and affecting the level of protection from harm they provide.[522]
- Our current evidence base was not sufficiently robust to demonstrate that available tools can reliably detect the harms in scope of the measures, while accounting for context. They raised these arguments in particular for grooming, eating disorder content, suicide content, self-harm content and pornography.[523]
- Our measures lacked clarity on what actions providers should take in respect of false positives, including what an "acceptable level" of false positives is and what redress and restoration processes there would be if content was wrongly taken down as a false positive.[524]

---

[520] Cook, J. response to the June 2025 Additional Safety Measures Consultation, p. 2; [✂]; [✂]; Open Rights Group response to the June 2025 Consultation, pp. 45-46; Professor Korff, D. response to the June 2025 Consultation, pp. 5, 8-9; Smith, G. response to the June 2025 Consultation, p. 5; [✂].
[521] Marie Collins Foundation response to the June 2025 Additional Safety Measures Consultation, p. 7; The Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p. 5.
[522] [✂]; Glitch response to the June 2025 Consultation, pp. 11-12; Open Rights Group response to the June 2025 Consultation, pp. 45-46; NSPCC response to the June 2025 Consultation, pp. 52-54; [✂].
[523] Antisemitism Policy Trust response to the June 2025 Additional Safety Measures Consultation, p. 2; Are, C, Dr. response to the June 2025 Additional Safety Measures Consultation, p. 5; Computer and Communications Industry Association's response to the June 2025 Consultation, p. 6; Cook, J. response to the June 2025 Consultation, pp. 1-2; European Digital Rights Organisation response to the June 2025 Additional Safety Measures Consultation, p. 1; F2F's response to the June 2025 Consultation, p. 2-3; GitHub's response to the June 2025 Additional Safety Measures Consultation, p. 2; Glitch response to the June 2025 Consultation, pp. 3-4, 6; Google LLC response to the June 2025 Consultation, pp. 5, 13-15; [✂]; ICO response to the June 2025 Consultation, p. 2, 8; International Justice Mission response to the June 2025 Additional Safety Measures Consultation, pp. 15-16; Microsoft response to the June 2025 Additional Safety Measures Consultation, p. 4; Name Withheld 4 response to the June 2025 Additional Safety Measures Consultation, p. 3; NetChoice response to the June 2025 Additional Safety Measures Consultation, pp. 1-2; NPCC response to the June 2025 Consultation, pp. 3-4; NSPCC response to the June 2025 Consultation, pp. 24-25; Open Rights Group response to the June 2025 Consultation, pp. 13-16, 18; Professor Korff, D. response to the June 2025 Consultation, pp. 4-5; Refuge response to the June 2025 Consultation, pp. 5-6; Respondents to the Together Declaration Campaign response to the June 2025 Additional Safety Measures Consultation, p. 5; Smith, G. response to the June 2025 Consultation, pp. 1-2; Social Media Animal Cruelty Coalition response to the June 2025 Additional Safety Measures Consultation, p. 8; The Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p. 4; Thompson, A. response to the June 2025 Additional Safety Measures Consultation, p. 3; [✂]; [✂]; [✂]; [✂]; [✂].
[524] Are, C. Dr response to the June 2025 Consultation, p. 15; BILETA response to the June 2025 Consultation, p. 8; Domestic Abuse Commissioner response to the June 2025 Additional Safety Measures Consultation, p. 5; Evangelical Alliance's response to the June 2025 Additional Safety Measures Consultation, p. 3; Glitch response to the June 2025 Consultation, p. 3; ICO response to the June 2025 Consultation, p. 10; Marie Collins Foundation response to the June 2025 Consultation, p. 7; Name Withheld 4 response to the June 2025 Consultation, pp. 3-4; Online Responsibility Network response to the June 2025 Additional Safety Measures Consultation, p. 3; Open Rights Group response to the June 2025 Consultation, pp. 14-15; Professor Korff, D. response to the June 2025 Consultation, pp. 11-12; Respondents to the Open Rights Group Campaign response to the June 2025 Consultation, p. 2; Smith, G. response to the June 2025 Consultation, p. 2; [✂].

- Our measures had a chilling effect and as a result of the measures, providers will choose to adopt more stringent terms of service and will moderate lawful content outside of the scope of the measure, or that users will choose to self-censor as a result of the measure.[525]
- Our measures should only apply to illegal content, and the inclusion of content harmful to children risks over blocking lawful dissent. Stakeholders argued Primary Priority Content is not well defined in law or guidance and imposes onerous measures on providers, which are not proportionate to identified risks and are not properly balanced with user rights.[526]
- Our measures risk enabling social or political control with speech limited to what is acceptable to those in power, and that proactive technology could be repurposed in repressive regimes, increasing risks for activists, journalists and human rights defenders.[527]
- The benefits of protecting users from the harms listed in the measures does not outweigh negative impacts on users' rights or justify interference with freedom of expression and association.[528]
- Our measures also raised various concerns about users' data protection and privacy rights, particularly the risk of excess data collection or surveillance of communications. We outline this feedback in more detail in the section entitled 'Rights impact'.

Our decision

Changes to the measures

9.11 In light of this feedback, we have decided to amend the measures we proposed in the consultation.

---

[525] BILETA response to the June 2025 Consultation, p. 8; Cook, J. response to the June 2025 Consultation, p. 2; European Digital Rights Organisation response to the June 2025 Consultation, p. 2; [✂]; Name Withheld 4 response to the June 2025 Consultation, p. 3; Name Withheld 22 response to the June 2025 Additional Safety Measures Consultation, p. 3; Professor Korff, D. response to the June 2025 Consultation, p. 6; Respondents to the Open Rights Group Campaign response to the June 2025 Consultation, p. 2; Respondents to the Together Declaration Campaign response to the June 2025 Consultation, p. 4-5; [✂]; The Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p. 5; [✂]; [✂].

[526] [✂]; Edwards, L. response to the June 2025 Additional Safety Measures Consultation, p. 3; [✂]; Google LLC response to the June 2025 Consultation, p. 4, 14; Name Withheld 4 response to the June 2025 Consultation, p. 3; NetChoice response to the June 2025 Consultation, p. 1; Respondents to the Open Rights Group Campaign response to the June 2025 Consultation, p. 2.

[527] Chayn response to the June 2025 Additional Safety Measures Consultation, p. 3; [✂]; Luethje, Y. response to the June 2025 Additional Safety Measures Consultation, p. 3; Mack, S, Dr. response to the June 2025 Additional Safety Measures Consultation, p. 3; Name Withheld 22 response to the June 2025 Consultation, p. 3; NetChoice response to the June 2025 Consultation, pp. 1-2; Newdick, P. response to the June 2025 Additional Safety Measures Consultation, p. 1; Open Rights Group response to the June 2025 Consultation, p. 13; Porter, J. response to the June 2025 Additional Safety Measures Consultation, p. 3; Respondents to the Together Declaration Campaign response to the June 2025 Consultation, pp. 4-5.

[528] Francis, S. response to the June 2025 Additional Safety Measures Consultation, pp. 7-8; Klaushofer, A. response to the June 2025 Additional Safety Measures Consultation, p. 4; Name Withheld 26 response to the June 2025 Additional Safety Measures Consultation, p. 5; Name Withheld 27 response to the June 2025 Additional Safety Measures Consultation, p. 5; [✂]; Respondents to the Together Declaration Campaign response to the June 2025 Consultation, p. 2; [✂]; [✂].

9.12 While we recognise the importance of giving providers flexibility to determine the type of proactive technology that is appropriate for their service, we also recognise stakeholder feedback that the measures would benefit from greater specificity.

9.13 We have therefore decided to recommend that service providers use classes and subclasses of proactive technology we consider are capable in principle of being accurate, effective and lacking in bias where they meet our criteria.[529] We refer to these classes and subclasses as 'specified technologies'. We refer to the particular technologies a provider selects from the list of specified technologies as 'relevant technology'. We are no longer recommending that providers assess whether proactive technology that meets our criteria exists. The specified technologies we are recommending in this measure are text matching and application-specific machine learning, and we define these in table 9.3.

9.14 These specified technologies are broad categories of proactive technology grouped according to their underlying technical approach or functionality, rather than specific products or vendors. We consider this approach maintains the degree of flexibility necessary for providers to address the way harms manifest on their service.[530]

9.15 In considering stakeholder feedback alongside our own analysis, we have a high degree of confidence in the specified technologies we have recommended for the purpose of detecting unknown image-based CSAM.[531] We are therefore confirming this amended version of the measure for image-based CSAM content. Therefore, this measure only applies to services at medium or high risk of image-based CSAM (see the section entitled 'Who the measure applies to' for more detail). We are revising our proposals with respect to other harms, and plan to consult in Spring 2027.[532]

9.16 We are also making the following amendments to our measures:

- Rather than two measures, we are now recommending one measure (which applies both to providers who do not have existing proactive technology for the purpose of detecting unknown image-based CSAM and those who do).
- We are not recommending that providers deploy proactive technology for CSAM proxy. We explain our reasoning for this change in the section entitled 'How the measure works'.

---

[529] We consider this addresses one stakeholder's recommendation that Ofcom share examples of technologies that comply with the proactive technology measures to illustrate good practice. NSPCC response to the June 2025 Consultation, p. 26.

[530] For these reasons we do not consider it is appropriate for us to certify specific tools or vendors, as recommended by some stakeholders. Image Angel response to the June 2025 Additional Safety Measures Consultation, pp. 5-6; Microsoft response to the June 2025 Consultation, p. 4. We note in the previous chapter that proactive technology and accredited technology are two distinct concepts under the Act, where accredited technology goes through a process of becoming accredited for use in a technology notice under section 121 of the Act. For more detail, see chapter 8: Introduction to Proactive Technology in the section entitled 'The distinction between proactive technology and accredited technology'.

[531] BILETA response to the June 2025 Consultation, p. 9; Common Sense Media's response to the June 2025 Consultation, p. 5; Evangelical Alliance's response to the June 2025 Consultation, p. 3; [✂]; [✂]; Online Safety Act Network response to the June 2025 Consultation, p. 19; Pinterest response to the June 2025 Consultation, p. 2.

[532] We detail the stakeholder feedback we received on the inclusion of additional harms to those we consulted on in the relevant section in annex 1: Further stakeholder responses.

- We have amended the information that providers should record as part of the measure. We explain our reasoning for this change in the section entitled 'Benefits and effectiveness'.

9.17 The full text of this measure can be found in Draft Illegal Content Code of Practice for User-to-User Services: Proactive Technologies (unknown content) and is referred to as 'Deployment of proactive technology to detect image-based CSAM content' (ICU PT1). We have created a single Code of Practice for this measure instead of putting it into our existing Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes). We intend to add more proactive technology measures (which may also list multiple specified technologies within one measure) to our Illegal Content Code: Proactive technology for further harms over time. This may include adding proactive technology measures related to both illegal harms and content harmful to children. It is also likely that we will amend or add to the specified technologies in each measure as technology evolves. We consider that putting these measures, and associated amendments, into one Code of Practice will allow providers to more easily access and understand how to apply them over time.

9.18 In response to stakeholders' feedback, we have also added a measure to our content moderation measures in the Illegal Content and Protection of Children Codes of Practice (for both user-to-user and search services) to ensure that decisions in respect of content or users are reversed, so far as possible and appropriate, where the provider finds them to be false positives. We explain our reasoning and outline the measures in the section titled 'Steps to take on reversal of content decisions'.

Changes to the proactive technology criteria and draft guidance

9.19 Stakeholder feedback on the rights impacts of the measures, outlined in the section entitled 'Summary of stakeholder feedback', led us to make amendments to both the proactive technology criteria and draft guidance.

9.20 Within the proactive technology criteria, we have made the following changes:

- Relocated the proactive technology criteria from the 'Key definitions' section to the 'Recommendation' section of the Code.
- Amended reference from the service's risk assessment to the provider's risk assessment in the criterion 'High-quality data'.
- Removed reference to the design and development process within the criterion 'Addressing biases'.
- Clarified the objective providers should aim for when determining an appropriate balance between precision and recall within the 'Evaluating performance' criterion.
- Included reference to 'evaluation' and 'assessing' within the 'Contextual testing and evaluation' criterion.
- Included references to quality assurance within the criteria 'Maintenance and ongoing monitoring', 'Human review' and 'Incorporating feedback'.
- Clarified within the 'Human review' criterion:
  - That the provider should abide by its policies and follow its processes for human review;
  - That human review serves two purposes, namely as part of moderation decisions and for quality assurance, and when human review should be conducted within the lifecycle of the proactive technology;

- The factors providers should take into account when determining the appropriate proportion of outputs subject to human review for both of these purposes; and
- That human review for quality assurance purposes will include the review of undetected content (i.e. suspected false negatives).

- Included reference to providers identifying, correcting and responding to errors and performance issues within the criterion 'Incorporating feedback'.

9.21 We explain our reasoning for the changes we have made to the criteria further in the section entitled 'Benefits and effectiveness'.

9.22 We have also made changes to the Proactive Technology Draft Guidance:

- The guidance now focuses on how providers should configure and deploy proactive technology, rather than how they should assess it. This is consistent with the changes we have made to the measures outlined in the section entitled 'Changes to the measures'.
- We have made a small amendment to the guidance describing our criterion on 'High-quality data' to clarify that trusted flagger reports may be included in the data that providers use to train proactive technology. We explain our reasoning for this amendment in more detail in the section entitled 'Benefits and effectiveness'.
- We have also made several amendments to the guidance to address feedback related to impacts on user rights. We explain our reasoning for these amendments in the section entitled 'Rights impact' and outline these in detail in annex 4.2: Proactive technology: draft guidance changes.

9.23 The proactive technology criteria are included in the full text of this measure in our Draft Illegal Content Code of Practice for User-to-User Services: Proactive Technologies (unknown content) (ICU PT1). The full guidance can be found in our regulatory documents published alongside this Statement and is titled 'Draft Proactive Technology Guidance'. We will issue the final Guidance when we issue the Codes.

How the measure works

Summary of the measure

9.24 We are recommending providers should use one (or a combination) of the specified technologies listed in table 9.3 to analyse relevant inputs for the purpose of detecting or supporting the detection of unknown image-based CSAM content. In selecting, and throughout the lifecycle of, the relevant technology concerned, the provider should ensure it meets the proactive technology criteria. We have assessed that in principle these specified technologies are capable of being accurate, effective and lacking in bias when providers do this.

Proactive technology criteria

Table 9.2 Proactive technology criteria

Criteria	Description
High-quality data	The high-quality data criterion is fulfilled if the relevant technology[533] has been developed and tested using high-quality[534] datasets appropriate to and reflecting a broad range of inputs relevant to the harm it is intended to detect (as identified in the provider's risk assessment).
Addressing biases	The addressing biases criterion is fulfilled if: a) Potential biases have been identified and addressed; and b) Risks are appropriately managed and addressed through the relevant technology's lifecycle.
Evaluating performance	The evaluating performance criterion is fulfilled if: a) The relevant technology has been evaluated using appropriate performance metrics; and b) Configured so that its performance strikes an appropriate balance between precision[535] and recall.[536]
Safeguards against misuse and exploitation	The safeguards against misuse and exploitation criterion is fulfilled if safeguards are in place to identify and appropriately manage security threats and risks of exploitation and misuse, including through the use of access restrictions[537] and system integrity protections.

[533] A form of proactive technology that constitutes one or a combination of the specified technologies for the purposes of ICU PT 1 and has been implemented by a provider to detect target content.

[534] The Draft Proactive Technology Guidance explains that "High quality" refers to data that is accurate; legally and ethically sourced; well-labelled (where appropriate); representative of the harm and context the proactive technology is intended to address; and sufficiently diverse to support meaningful evaluation, including (where relevant) to test the proactive technology's ability to detect content it has not previously encountered. The degree to which a dataset may be considered "high-quality" will vary based on the harm being addressed and the context in which the proactive technology is used. However, indicators that a dataset might not be considered "high quality" would include (for example) unclear sourcing, evidence of poor or inconsistent labelling, or limited representation of relevant harms or input types.

[535] Precision is a measure of statistical accuracy, calculated as the proportion of detected content that a relevant technology has correctly identified as target content.

[536] Recall is a measure of statistical accuracy, calculated as the proportion of target content analysed by a relevant technology that the technology has detected.

[537] The Draft Proactive Technology Guidance explains that "Access restrictions" refers to limitations placed on interaction with system operations or data, based on the roles or responsibilities assigned to any entity (including individuals, devices, or systems such as software components, applications, and automated processes), whether internal and external. This aligns with principles on controlling access, as outlined in the UK Cyber Assessment Framework, and supported by international standards such as ISO/IEC27001 and NIST SP-800-53.

Criteria	Description
Contextual testing and evaluation	The contextual testing and evaluation criterion is fulfilled if: a) The relevant technology's performance has been evaluated in real-world use cases relevant to the provider's content (having regard to this risk of harm to individuals identified in the provider's risk assessment); and b) The results indicate that it correctly detects the harm it is intended to detect.
Maintenance and ongoing monitoring	The maintenance and ongoing monitoring criterion is fulfilled if mechanisms are in place to monitor, maintain and quality assure the relevant technology's effectiveness over time, including processes for regular review, quality assurance and iterative adjustments to respond to emerging circumvention techniques, biases, or new content types.
Human review	The human review criterion is fulfilled if: a) Policies and processes are in place for human review for both moderation and quality assurance purposes; and b) The provider abides by its policies and follows its processes.
Incorporating feedback	The incorporating feedback criterion is fulfilled if feedback and quality assurance mechanisms are in place to maintain or improve performance of the relevant technology and associated systems and processes over time. This includes: a) Identifying, correcting and responding to errors and performance issues; b) Updating the relevant technology with diverse and up-to-date datasets to reflect evolving trends or emerging types of image-based CSAM content; and c) Integrating ongoing feedback from users and individuals working in moderation[538] into its development, while managing the risk of introducing additional bias.

Additional factors to consider for 'Evaluating performance'

9.25 For the criterion 'Evaluating performance', providers should consider certain factors related to the testing, configuration, deployment and ongoing monitoring of proactive technology.

9.26 An appropriate balance of precision and recall is a balance set with a view to minimising the incidence and impact of moderation errors including both failures to identify image-based CSAM content and incorrect action against content. When configuring the technology so that it strikes an appropriate balance between precision and recall, the provider should ensure that the following factors are taken into account:

a) The service's risk of image-based CSAM, reflecting the risk assessment of the service and any information reasonably available to the provider about the prevalence of target content on the service;
b) The proportion of detected content that is a false positive;

[538] For an explanation of 'individuals working in content moderation', see our December 2024 Statement: Volume 2: Service design and user choice, pp. 79-80.

c) The effectiveness of the systems and/or processes used to identify false positives; and
d) The importance of minimising the reporting of false positives to the National Crime Agency (NCA) or a foreign agency.

Additional factors to consider for 'Human review'

9.27 For the criterion 'Human review', providers must:

a) For moderation purposes, ensure the human review of an appropriate proportion of the outputs of the relevant technology during deployment;
b) For quality assurance purposes, throughout the lifecycle of the relevant technology, ensure the human review of an appropriate proportion of the outputs of the proactive technology and of undetected content;
c) Ensure outputs are explainable to the extent necessary to support meaningful human judgement and accountability; and
d) When determining what is an appropriate proportion of outputs of the relevant technology and undetected content (where applicable) to be reviewed by humans, ensure that the following factors are taken into account:

i) The level of assurance the provider has in the detection outcomes produced by the technology and any associated systems and processes;
ii) The potential severity of the harm to those depicted in or encountering image-based CSAM content; and
iii) For moderation purposes, the overall impact of an incorrect decision that the content is image-based CSAM content, on the user who generated, uploaded or shared the content, having particular regard to the importance of minimising the reporting of false positives to the National Crime Agency (NCA) or a foreign agency.

9.28 Further guidance on the proactive technology criteria, including examples of how to configure and deploy proactive technology against the criteria and how they can be met, are set out in the Draft Guidance on Proactive Technology. This was published alongside this statement.

Using proactive technology

9.29 We have assessed the specified technologies presented in the following table and consider them to be capable in principle of being sufficiently accurate, effective and lacking in bias if configured and deployed in accordance with the proactive technology criteria.[539] We set our reasoning for why we have recommended these specified technologies in annex 4.1 Proactive technology: CSAM proactive technology technical annex.

9.30 We therefore recommend that providers use one (or a combination)[540] of the specified technologies listed in table 9.3 for the purpose of detecting unknown image-based CSAM.

---

[539] We consider this addresses concerns or requests for clarification raised by stakeholders regarding our proposal in the June 2025 Consultation that providers assess technology themselves, as we have assessed these classes of technology. Middle Tech Coalition asked for clarity regarding whether these assessments could be consolidated and incorporated into existing risk assessment processes to avoid unnecessary duplication and additional compliance burdens. CCO response to the June 2025 Consultation, pp. 10-11; Glitch response to the June 2025 Consultation, p. 5,12; Meta response to the June 2025 Consultation, p. 9. Middle Tech Coalition response to the June 2025 Consultation, p. 3.

[540] The Internet Watch Foundation highlighted the benefits of service providers using a layered approach to detect CSAM, combining hash-matching, AI tools, and human moderation. We consider we have addressed

Table 9.3: Recommended specified technologies for the purpose of detecting unknown image-based CSAM

Specified technologies	Definition
Text matching (for the detection of images associated with the matched text)	Exact matching, approximate matching, or pattern matching technologies based on rule-based or heuristic matching approaches to identify predefined textual elements and patterns within content, including words, phrases, character sequences, uniform resource locators (URLs), and domain names.
Application-specific machine learning technologies	Technologies based on models developed or adapted using training data to identify, classify, or assess features, attributes, or characteristics of target content, and whose intended tasks, decision criteria, and expected outputs are established through model development or adaptation.

9.31 Image-based CSAM may manifest in a variety of ways across different services. Therefore, selecting the specified technologies (or a combination of specified technologies) for the purpose of detecting or supporting the detection of unknown image-based CSAM, providers should have regard to how image-based CSAM content manifests on their service as set out in their risk assessment.

9.32 In some cases, text linked to an image (for example comments or tags) may be an indicator of whether the image is CSAM. We are recommending that, in having regard to how image-based CSAM manifests on their service, providers should consider the particular needs of their UK user base in relation to languages. For more detail, see our Draft Guidance on Proactive Technology and annex 4.1: Proactive technology: CSAM proactive technology technical annex.

9.33 We are not recommending providers use proactive technology to detect CSAM content proxy in our Code measure.[541] This is because Ofcom may only recommend a proactive technology measure for the purpose of compliance with any of the duties regarding content as defined in the Act.[542] While it is at our discretion to assess and recommend specified technologies, for the detection of image-based CSAM, the Act does not require us to do so regarding proxy content.[543] Providers who prohibit the presence of CSAM content proxy on their service in their terms of service, and wish to use proactive technology to detect it, may do so as an alternative measure to those recommended in our Code.

9.34 Relatedly, two industry stakeholders said that proactive technology had to be developed globally, as building classifiers for hundreds of local laws was not feasible or proportionate.

---

this concern by allowing providers to use a combination of the classes of technology listed in the measure. Internet Watch Foundation response to the June 2025 Consultation, pp. 19-21.

[541] In our Illegal Content Code: Proactive technology, we define CSAM content proxy as content that a provider determines to be in breach of its terms of service, where: a) the provider had reason to suspect that the content may be CSAM; and b) the provider is satisfied that its terms of service prohibit the type of CSAM which it had reason to suspect existed.

[542] Paragraph 13(3) of Schedule 4 to the Act (namely sections 10(2), (3) or (3A), sections 12(2) or (3), sections 27(2), (3) or (3A), sections 29(2) or (3), or sections 38(1) or 39(1) of the Act).

[543] Our legal framework in annex 3 sets out definitions for content that is harmful to children proxy, CSAM content proxy, CSEA content proxy and illegal content proxy.

They warned that Ofcom's mandate could force providers to either apply UK standards globally, develop less effective UK-specific technology or fragment otherwise global services.[544] Providers are free to develop proactive technology to be applied globally, however they should be able to explain how the relevant safety duties in the Act have been met by doing so.

Using proactive technology for the purpose of detecting unknown image-based CSAM

9.35 Service providers should use, where technically feasible, specified technologies (or a combination of specified technologies) in table 9.3 for detecting unknown image-based CSAM on their service(s). In selecting, and throughout the lifecycle of, the relevant technology concerned, the provider should ensure that it meets the proactive technology criteria. We outline these in the section entitled 'Proactive technology criteria'.

9.36 Where a provider already uses one (or a combination) of the specified technologies in table 9.3, it should ensure that this existing proactive technology, throughout its lifecycle, meets the proactive technology criteria.[545] Where this cannot be achieved through the use of existing proactive technology, we consider proactive technology is available either sourced from a third party or through in-house development, or a combination of both.[546] [547] We set this out in more detail in annex 4.1: Proactive technology: CSAM proactive technology technical annex.

9.37 Whichever option a provider decides to adopt, we note that Codes measures, including this one, are a safe harbour. Service providers will not be in the safe harbour until they have configured and deployed proactive technology in accordance with the measure. We recognise that the work of deciding which technology to use, including carrying out appropriate tests, may take some time. In determining whether or not to enforce, we will have regard to whether a provider has acted within a reasonable timeframe. A reasonable timeframe for a provider to configure and deploy proactive technology will vary depending on several factors, including: the type of service and its functionalities; whether or not providers choose to develop technology in-house; and whether providers have an existing framework to integrate technology.[548] [549]

---

[544] Computer and Communications Industry Association's response to the June 2025 Consultation, p. 7; Google LLC response to the June 2025 Consultation, p. 5, 15.

[545] We consider this addresses concerns from providers about their ability to continue to use existing proactive technology they have developed in-house. Meta response to the June 2025 Consultation, p.10; [✂]; [✂]; [✂]. Where a provider wishes to use existing proactive technology that is not one of the recommended technologies listed in table 9.3, it should pursue an alternative measure to those set out in our Codes to meet its safety duties in the Act.

[546] We consider this addresses concerns about providers "opting out" of deploying proactive technology and the measure not recommending that providers without existing proactive technology should develop it. Digital Resilience in Education – Welsh Government response to the June 2025 Consultation, p. 6; UK Finance response to the June 2025 Consultation, p. 2, 7.

[547] We consider that this will satisfy concerns expressed by some stakeholders about providers using proactive technology that does not meet the criteria. ICO response to the June 2025 Consultation, p. 11; Integrity Institute response to the June 2025 Additional Safety Measures Consultation, p. 3; Name Withheld 4 response to the June 2025 Consultation, p. 11.

[548] [✂].

[549] We consider this addresses concerns that our requirement for platforms to act 'within a reasonable timeframe' was too vague and risked long delays. Glitch response to the June 2025 Consultation, p. 6.

9.38 Ofcom:

a) Has the power to investigate whether such providers are complying with their duties under Section 10 of the Act;
b) Recognises the seriousness of the harm this measure seeks to address; and
c) Accordingly expects providers of in scope services wishing to be in the safe harbour to take immediate steps to fully implement the measure promptly.[550]

9.39 We will act reasonably in exercising our powers where service providers are not able to fully deploy proactive technology immediately but demonstrate they are taking steps to implement the measure promptly following the Code coming into force. Where a provider has decided to develop proactive technology in-house, we may request to see its 'implementation plan' and its reasoning as to why it considers the timelines set out in this plan are reasonable. See the section entitled 'Recording and reviewing classes of technology' in the draft Proactive Technology Guidance for more detail.

Scanning content or inputs

9.40 When deploying proactive technology, we recommend providers should:

- Ensure that relevant inputs[551] that are generated on, uploaded to or shared on the service (or that a user seeks to so generate, upload or share) after the relevant technology is implemented are analysed before, or as soon as practicable after, relevant content[552] can be encountered by United Kingdom users of the service.
- Ensure that all relevant content present on the service when the technology is first implemented is analysed within a reasonable time.[553]

9.41 When deploying text matching technologies, the provider should treat detection by the technology as reason to suspect that items of content associated with the matched text may be illegal content and review that content in accordance with Recommendation ICU C1 of the Illegal Content Codes.[554]

9.42 When deploying application-specific machine learning technologies the provider should treat detection by the technology as reason to suspect that the content may be illegal content and review the content in accordance with Recommendation ICU C1 of the Illegal content Codes.[555]

9.43 Following review, a provider may determine that content detected to be image-based CSAM amounts to another priority offence in the Act. Where a provider makes this

[550] We consider that this will satisfy concerns expressed by some stakeholders about providers using proactive technology that does not meet the criteria. ICO response to the June 2025 Additional Safety Measures Consultation, p. 11; Integrity Institute response to the June 2025 Additional Safety Measures Consultation, p. 3; Name Withheld 4 response to the June 2025 Additional Safety Measures Consultation, p. 11.
[551] Relevant content or metadata relating to relevant content.
[552] Any: a) regulated user-generated content that: i) may be encountered by United Kingdom users of the service by means of the service, and ii) is communicated publicly by means of the service, or b) material which, if it were present on the service, would be content within paragraph (a).
[553] Two stakeholders supported our proposal that providers at risk of CSEA should scan all content, including pre-existing content. Canadian Centre for Child Protection response to the June 2025 Consultation, p. 9; [✂].
[554] These can be found written in full in our Illegal Content Codes.
[555] These can be found written in full in our Illegal Content Codes.

determination, they should take appropriate action in line with existing Illegal Content Codes.

9.44 We have outlined factors to assist providers in determining the appropriate proportion of detected content that should be subject to human review in content moderation in the section entitled 'Proactive technology criteria'.

9.45 Several stakeholders suggested that clarification was needed on how our measures would apply to livestreamed content, although some suggested the protection of users during livestreaming required additional interventions including audience controls, age assurance, human moderation, and rapid content removal.[556] Several stakeholders also highlighted the risk of CSAM on livestreaming services.[557] Some said that proactive technology was being used to detect harmful content, including CSAM, on services with livestreaming functionality.[558]

9.46 As outlined in annex 4.1: Proactive technology: CSAM proactive technology technical annex, the specified technologies we are recommending are capable of detecting livestreamed content and therefore livestreamed content is covered by the measure.[559]

Recording and reviewing

9.47 The provider should maintain sufficient records to demonstrate the basis on which it concluded the relevant technology met the proactive technology criteria, including:

a) Records of any testing or evaluation undertaken,
b) The performance metrics used,
c) The results obtained,
d) Configuration decisions,
e) Identified limitations and mitigations, and
f) Any ongoing monitoring and review activities relevant to its assessment.

9.48 At least every 12 months, or earlier where there is a change to any aspect of the service's design or operation which would amount to a significant change and therefore trigger a risk assessment under section 9(4) of the Act, the provider should review its selection of relevant technology.

---

[556] A member of the Philippine Survivor Network response to the June 2025 Consultation, pp. 7-13; Antisemitism Policy Trust response to the June 2025 Consultation, p. 1; Barnardo's response to the June 2025 Consultation, p. 5; BILETA response to the June 2025 Consultation, p. 2; Canadian Centre for Child Protection response to the June 2025 Consultation, p. 11; CCO response to the June 2025 Consultation, pp. 10-11; Computer and Communications Industry Association response to the June 2025 Consultation, pp. 4-5; International Justice Mission response to the June 2025 Consultation, pp. 8-13; Internet Watch Foundation response to the June 2025 Consultation, pp. 2-3; Marie Collins Foundation response to the June 2025 Consultation, p. 8; South West Grid for Learning response to the June 2025 Consultation, p. 9; The Domestic Abuse Commissioner response to the June 2025 Consultation, pp. 4-5.

[557] International Justice Mission response to the June 2025 Consultation, pp. 8-13; Internet Watch Foundation response to the June 2025 Consultation, pp. 2-3; South West Grid for Learning response to the June 2025 Consultation, p. 9.

[558] A member of the Philippine Survivor Network response to the June 2025 Consultation, pp. 8-9; Computer and Communications Industry Association response to the June 2025 Consultation, pp. 4-5; [✂]; [✂]; [✂].

[559] In response to several stakeholders who requested clarification, we can confirm that content generated using AI should be captured by the measures in the same way as human-generated content. BILETA response to the June 2025 Consultation, p. 2; Common Sense Media response to the June 2025 Consultation, pp. 6-7; NSPCC response to the June 2025 Consultation, pp. 31-32.

9.49 Providers should also keep the records required by the record-keeping duty, in line with Ofcom's Record-keeping and Review Guidance.[560]

9.50 We consider that if a provider's existing proactive technology cannot be deployed and configured in a way that meets the criteria, proactive technology is available either sourced from a third party or through in-house development, or a combination of both as outlined in the section entitled 'Using proactive technology for the purpose of detecting unknown image-based CSAM'. However, where a provider has not deployed any proactive technology because it is not technically feasible to do so, we would expect it to record this decision and its reasoning.[561] As outlined in the section entitled 'Technical feasibility' in chapter 8: Introduction to proactive technology, where a service provider claims that a measure is technically infeasible, Ofcom has the power to investigate.

Our reasoning

Benefits and effectiveness of the measure

9.51 In this section, we consider the benefits of using proactive technology to increase the safety of UK users and outline its effectiveness at detecting unknown image-based CSAM. We first outline:

- The risk that this measure seeks to address; and
- Why proactive technology is necessary to address unknown image-based CSAM;

9.52 We then outline the benefits and effectiveness of specific recommendations that we have made as parts of our measure, including:

- The accuracy, effectiveness and lack of bias of our specified technologies;
- What providers should analyse;
- Recording and reviewing the measure; and
- The proactive technology criteria and draft guidance.

The risk that this measure seeks to address

9.53 The risk of online image-based CSAM is the risk of an offence specified in any of paragraphs 1 to 4, 7, 8 or 10 of Schedule 6 to the Act,[562] so far as the risk in relation to these offences relates to CSAM in the form of photographs, videos or visual images. It includes material depicting sexual activity involving children, or indecent or prohibited imagery of children, and can take the form of photographic images and videos, as well as non-photographic material such as drawings and animations. It can also be content generated by artificial intelligence (AI).[563] The impacts of this harm, which include both physical and psychological impacts, are widespread and devastating, with some victims affected for the rest of their

[560] The written record should be durable, accessible, easy to understand and up to date. Ofcom (2024), Record-Keeping and Review Guidance.

[561] We consider that this addressed feedback from one stakeholder about whether a record needs to be made where a service has concluded that it is not technically feasible to implement proactive technology. techUK response to the June 2025 Consultation, p. 12.

[562] See Table C at paragraph 5.6 in the Illegal Content Codes.

[563] The NSPCC said that a wider range of research that generative-AI services can be used to create illegal and harmful content, and also act as a source of harmful content to children. NSPCC response to the June 2025 Consultation, pp. 30-31.

lives. We outline the severity and impact of this harm below, and in further detail in chapter 2: child sexual exploitation and abuse (CSEA) of the Illegal Harms Register of Risks.[564]

9.54 CSAM can have a profound and long-lasting impact on children who are sexually abused, as well as on the wellbeing of adults and children who encounter this material. Beyond the abuse itself, the presence, sharing and viewing of images and videos depicting the abuse can serve as a continual source of trauma for victims and survivors of CSEA.[565] This can include reliving their abuse if they encounter the material online, or the fear many victims experience of being recognised by someone who has seen it.

9.55 CSAM is often accessed intentionally, and the availability of such material online creates a permissive environment in which perpetrators may develop and act on their sexual interests.[566] The availability of CSAM can also lead to unintentional viewing, likely causing considerable distress.[567] For example, children who had accidentally accessed CSAM online often felt reluctant to confide in anyone about it, fearing they might not be believed or that they could face arrest.[568]

9.56 Children themselves may generate content that meets the definition of CSAM, which can cause them harm.[569] Children may do so as part of a consensual peer-to-peer relationship or following coercion or grooming. Once shared by the child, these images or videos may be shared more widely without the child's consent or used to blackmail the child into creating and sharing further sexual images.

9.57 The scale of online CSAM has been increasing year-on-year. In 2025, the Internet Watch Foundation (IWF) reported a 7% increase compared with 2024 in the number of reports confirmed as or linked to CSAM, rising to over 310,000 confirmed reports.[570] In the same period, the IWF has also seen a 50% increase in the number of child sexual abuse videos confirmed as reaching the threshold of criminality, totalling over 63,000 confirmed reports.[571] This included a 29% increase in material representing the most severe form of CSAM (Category A).[572] [573] Further, the IWF has also reported a 26,385% increase in AI-

[564] We set out the evidence for the severity of harms and the impact on victims and survivors in full in the relevant chapters of our Illegal Harms Register of Risks.
[565] Hamilton-Giachritsis, C., Hanson, E., Whittle, H. and Beech, A., 2017. "Everyone deserves to be happy and safe": A mixed methods study exploring how online and offline child sexual abuse impact young people and how professionals respond to it.
[566] Insoll, T., Ovaska, A. K., Nurmi, J., Aaltonen, M., and Vaaranen-Valkonen, N., 2022. Risk Factors for Child Sexual Abuse Material Users Contacting Children Online: Results of an anonymous multilingual survey on the dark web, Journal of Online Trust and Safety, 1(2).
[567] Ofcom's Online Experiences Tracker found that in H2 2026, 4% of adults had encountered content depicting the sexual abuse or sexual exploitation of children in the past 4 weeks.
[568] NSPCC, 2016. Online child sexual abuse images: Doing more to tackle demand and supply.
[569] The Internet Watch Foundation found that, of the 56,179 child sexual abuse images it analysed, one-third were 'self-generated'. Internet Watch Foundation, 2025. Executive Summary, IWF Annual Data & Insights Report 2025, p. 14.
[570] Internet Watch Foundation, 2025. Executive Summary, IWF Annual Data & Insights Report 2025.
[571] Internet Watch Foundation, 2025. Executive Summary, IWF Annual Data & Insights Report 2025.
[572] His Majesty's Inspectorate of Constabulary and Fire & Rescue Services (HMICFRS) (n.d.) Child sexual abuse image grading.
[573] Internet Watch Foundation, 2025. Executive Summary, IWF Annual Data & Insights Report 2025.

generated CSAM videos from 2024 to 2025.[574] [575] These figures are based on reports to the IWF from the public, industry, law enforcement and IWF partners, and therefore represent only detected and reported CSAM. The true scale of online CSAM content is likely to be significantly higher. The NSPCC also reported that in 2025, 36,775 recorded offences of indecent and prohibited images of children were recorded across 41 different police forces, representing an 8% rise compared to 2024.[576]

9.58 We have already recommended that providers use hash matching to detect and remove known CSAM (ICU C9). As explained in chapter 8: Introduction to proactive technology, hash matching refers to a process for detecting content that has previously been identified as being illegal, or otherwise violative, to prevent the upload and sharing of that content.

9.59 We consider that there is an additional substantial risk to UK users posed by unknown image-based CSAM, which is CSAM content that is not included in any hash list used by the service provider. The NSPCC and Barnardo's both noted that only a fraction of CSAM online is known and that existing hash-matching recommendations, while important, do not address unknown image-based CSAM.[577] The growing availability and use of tools to create AI-generated CSAM is also making it easier, cheaper and quicker for offenders to produce and share unknown CSAM. Further, it takes time for CSAM images to appear on hash lists, and in light of the evolving harm from AI-generated CSAM we consider that there is a need for services to do more to detect CSAM that does not currently appear on hash lists.

9.60 Proactive technology that is designed and deployed to detect unknown CSAM is also capable of detecting known CSAM. However, to ensure the detection and removal of CSAM is as quick and effective as possible, we consider it is proportionate that in-scope services continue to use our recommended measure on hash matching for known CSAM (ICU C9) alongside proactive technology for unknown image-based CSAM. Hash matching is a well-established industry practice that is highly accurate at detecting known CSAM and enables automated removal and reporting of that content without further layers of moderation or human review. Hash matching is also necessary to protect child users who may appear to be older than they are, as review of the content may not be sufficient on its own to identify it as CSAM. Proactive technology is vital for the detection of unknown CSAM but should be implemented alongside hash matching as it may not always detect known content with the same level of accuracy or confidence as hash matching.

9.61 Therefore, due to the severity of unknown image-based CSAM and its prevalence online, we consider proactive measures are needed to prevent UK users from encountering it. We set out why we consider proactive technology can help mitigate the spread of unknown image-based CSAM in more detail in the following paragraphs.

Why proactive technology is necessary to address unknown image-based CSAM

9.62 We consider that deploying proactive technology will materially enhance providers' ability to detect (and swiftly take action against) image-based CSAM. Further, we consider that the

---

[574] Internet Watch Foundation, 2026. AI CSAM Report 2026 – Harm without limits: AI child sexual abuse material through the eyes of our Analysts.

[575] This report was also referenced by the NSPCC, who stated our proactive technology measures were important for tackling the rise in AI-generated CSAM. NSPCC response to the June 2025 Consultation, p. 20.

[576] NSPCC, 2026. We're calling for tech companies to block child sexual abuse images in real time.

[577] Barnardo's response to the June 2025 Consultation, p. 7; NSPCC response to the June 2025 Consultation, p. 20.

use of these tools will be necessary for providers to effectively combat unknown image-based CSAM where they do not already implement them.

9.63 A range of stakeholders expressed their support for the use of proactive technology by service providers more generally, including to detect illegal content or content harmful to children,[578] and as a means to prevent the spread of CSAM.[579]

9.64 There are several reasons why we agree with stakeholders that the use of proactive technology is necessary to address unknown image-based CSAM, which includes but is not limited to:

- **The volume of content**: Given the high volume of content that is shared, uploaded and generated on user-to-user services, proactive technology is needed to detect the presence of suspected image-based CSAM among this content. For large services in particular, we understand the volume of content uploaded may range from millions to billions of pieces of content every day.[580]
- **The speed at which content spreads**: Features and functionalities that allow users to share content with one another (direct messaging) and amplify content (recommender systems) can result in content spreading quickly. This means that image-based CSAM may be uploaded, generated and shared at speed to a large number of UK users before it is detected. We consider proactive technology is necessary to detect image-based CSAM before UK users can encounter this content, or before it has the opportunity to spread.
- **The nature of the internet**: The internet enables offenders to connect and share image-based CSAM internationally. Offenders can easily find other offenders with relative anonymity and across borders.
- **Limitations of solely relying on user reporting**: While user reporting and complaints systems are important in detecting illegal content and as part of a wider safety strategy, we do not consider this alone is sufficient to mitigate the risk of image-based CSAM. Further, we do not consider it is sufficient to address the scale and volume of content uploaded to services in a way that prevents users from encountering this content.[581] Evidence suggests that some

[578] 5Rights Foundation response to the June 2025 Additional Safety Measures Consultation, p. 9; Department of Justice for Northern Ireland response to the June 2025 Additional Safety Measures Consultation, p. 7; Digital Resilience in Education - Welsh Government response to the June 2025 Consultation, p. 6; [✂]; Microsoft response to the June 2025 Consultation, p. 3; Molly Rose Foundation response to the June 2025 Consultation, p. 15; Name Withheld 11 response to the June 2025 Additional Safety Measures Consultation, p. 2; UK Finance response to the June 2025 Consultation, p. 2.

[579] A member of the Philippine Survivor Network response to the June 2025 Consultation, pp. 10-11; BILETA response to the June 2025 Consultation, p. 8; Canadian Centre for Child Protection response to the June 2025 Consultation, p. 9; Evangelical Alliance response to the June 2025 Consultation, p. 3; International Justice Mission response to the June 2025 Consultation, p. 10; The Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, p. 4; IWF response to the June 2025 Consultation, p. 13.

[580] Publicly available figures from providers of large services. For example, Reddit reported that, globally, approximately 260 million posts and 2 billion comments were created in H2 2025, equivalent to around 1.4 million posts and 11 million comments per day. Reddit, 2026. Transparency Report; July to December 2025. Snap reported that in 2025, global users on average posted 5.5 billion Snaps per day. Snap Inc, 2026. Snapchatters Created Nearly 2 Trillion Snaps in 2025.

[581] One stakeholder noted that the volume of CSAM online means that user reporting and human moderation alone cannot be relied upon to detect and swiftly remove it. Lucy Faithfull Foundation response to the June 2025 Additional Safety Measures Consultation, p. 4.

users perceived user reporting to sometimes be ineffective at taking down illegal content and/or content harmful to children.[582] [583] It is possible that, as a result of this, users may be deterred from reporting as they consider it is unclear what happens after submitting a report.[584] While we have recommended measures to improve user reporting processes, we consider that proactive technology is needed on top of this to ensure image-based CSAM is detected where users are not reporting it. Evidence also suggests that, particularly for children, there is a reluctance to report self-generated indecent images (SGII) due to shame, fear, or belief that nothing can be done and not recognising that abuse is occurring (for example believing they are in a relationship with an adult).[585] Furthermore, CSAM is often shared among offenders, and so none of those involved will report the content.

- **Limitations of manual human detection**: Proactive technology can analyse content and associated inputs at a scale and speed that cannot practicably be achieved through manual review alone. Depending on the technology, it may also combine multiple inputs to support detection, including inputs relevant to estimating whether a depicted individual may be a child.

9.65 Therefore, given the volume and scale of user-generated content that is uploaded, generated and shared to a service, the limitations of relying solely on user reporting and manual human detection, and the nature of the internet, we consider that proactive technology is necessary to detect unknown image-based CSAM. We consider this will lead to an increase in the detection and removal of CSAM on services and reduce the amount of image-based CSAM in circulation more widely.

Effectiveness, accuracy and lack of bias of our specified technologies

9.66 In this section, we set out how our specified technologies can help protect users from encountering image-based CSAM. In particular, we explain why we consider our specified technologies are capable in principle of detecting unknown image-based CSAM in a way that is accurate, effective and lacking in bias where they meet our proactive technology criteria.[586] For the purpose of this measure, we are recommending that providers use one (or a combination) of the specified technologies we set out in table 9.3. We outline this below, and in greater detail in annex 4.1: Proactive technology: CSAM proactive technology technical annex.

9.67 We note that we have considered technologies to detect the main kind of image-based CSAM, i.e. images of children. We have not assessed the use of the technology to detect images of text (for example, a photograph of a paedophile manual). For this reason, we

[582] YouGov & Ofcom, 2026. Exploring how internet users navigate sensitive and harmful content. We note the sample for this research was purposefully skewed to participants with minority identities that are more likely to be targeted in hateful content.
[583] Digital Resilience in Education – Welsh Government raised the additional concern that even if a user engages with content to report it, this is still considered engagement, and users may be shown more similar content. It stated its youth group advised it does not report content on some services because of this. Digital Resilience in Education – Welsh Government response to the June 2025 Consultation, pp. 6-7.
[584] Ofcom (2025), Behavioural Audit of Online Services.
[585] 5,862 victims and survivors participated in the Truth Project. The IICSA, 2022. The Report of the Independent Inquiry into Child Sexual Abuse.
[586] We set out the benefits of our proactive technology criteria in the section entitled 'Benefits and effectiveness of the proactive technology criteria and guidance'.

have defined the content providers should seek to detect in a way which focuses on images of children.[587]

9.68 We consider that text matching and application-specific machine learning technologies are each capable in principle of accurately and effectively supporting the detection of image-based CSAM, in a manner that lacks undue bias. While each technology class exhibits distinct strengths, limitations and deployment considerations, the available evidence indicates that those limitations are capable of mitigation where they meet the proactive technology criteria. We set this out in detail in annex 4.1: Proactive technology: CSAM proactive technology technical annex.

Application-specific machine learning technologies

9.69 Application-specific machine learning technologies are capable of directly analysing image and video content and identifying visual features associated with image-based CSAM. Where representative training data is available, such technologies are capable of achieving high levels of classification performance across image-classification tasks. While access to such datasets can present challenges in the case of CSAM, specialist organisations such as NCMEC have facilitated access to representative CSAM datasets and related training resources. We also consider that the limitations associated with these approaches can be mitigated through appropriate dataset governance, calibration, testing, monitoring and maintenance practices.

9.70 Taking these factors together, we conclude that application-specific machine learning technologies are capable in principle of accurately and effectively supporting detection of image-based CSAM, in a manner that lacks undue bias, where they meet the proactive technology criteria.

Text matching

9.71 We have found that textual indicators associated with image-based CSAM are sufficiently associated with the target harm to support meaningful detection outcomes and that such indicators can be operationalised through appropriately configured text matching systems. We have also found that identified limitations may be mitigated through the use of appropriate indicator datasets, ongoing maintenance and deployment practices consistent with the proactive technology criteria. Our analysis shows that a number of high quality and well-maintained datasets containing these textual indicators are available.

9.72 Taking these factors into account, we conclude that text matching technologies are capable in principle of being sufficiently accurate, effective and lacking in bias for the purpose of detecting unknown image-based CSAM where they meet the proactive technology criteria.

9.73 For these reasons, we consider that users will benefit significantly if providers use one or a combination of the specified technologies in table 9.3 for the purpose of detecting unknown image-based CSAM and swiftly take it down once it has been determined to be

[587] We have defined "image based CSAM content" for the purposes of the measure as "content amounting to an offence specified in any of the following paragraphs of Schedule 6 to the Act: 2 to 4, 7 or 10, so far as the risk in relation to those offences relates to CSAM in the form of photographs, videos or visual images". This omits the offences to do with obscene publications relating to child abuse and paedophile manuals. We do not consider it likely that a service provider would be at risk of hosting (for example) photographs of obscene publications relating to child abuse or paedophile manuals and not be at risk of hosting illegal images of children.

illegal content by providers' content moderation functions. We consider this contributes to preventing UK users from encountering image-based CSAM on their service.

What providers should analyse

9.74 We know that CSAM may circulate for years after it is first uploaded.[588] The repeated circulation of this content is a source of continued traumatisation and re-victimisation and is not lessened by how old it is. For these reasons, we recommend that providers should scan all content, including pre-existing content, on their service.[589]

9.75 The measure also sets out that providers should ensure that relevant inputs are analysed before, or as soon as practicable after, relevant content can be encountered by users after the relevant technology is implemented. We consider this provides significant benefits to user safety, as this means the time that image-based CSAM remains on the service and the number of UK users who encounter this content will be minimised.

Recording and reviewing

9.76 We consider that there are significant benefits of providers recording and reviewing the points outlined in the sub-section entitled 'Recording and reviewing' within the 'How the measure works' section. We also recommend providers review their selection of specified technologies every 12 months or earlier where there is a significant change to any aspect of the service design or operation and therefore triggers a risk assessment. We note several benefits to this approach:

i) First, we understand that the way image-based CSAM manifests on the service may change over time. This may affect which specified technologies are most appropriate for detecting unknown image-based CSAM. Where providers review the classes of proactive technology they have chosen to deploy on their service, they will be able to determine whether that technology remains appropriate for detecting unknown image-based CSAM.[590] [591] [592]

ii) Second, by recording whether the relevant technology concerned meets the proactive technology criteria, providers can make the necessary changes to ensure

---

[588] For example, the National Centre for Missing and Exploited Children (NCMEC) state in their 2024 report that CSAM of one child "has been circulated for the past 19 years appearing more than 1.3 million times in submissions" to the organisation. NCMEC, 2024. NCMEC 2024 CyberTipline Report.

[589] We consider that this addresses feedback from two stakeholders who expressed their support for our proposal that providers at risk of CSEA should scan all content, including pre-existing content. Canadian Centre for Child Protection response to the June 2025 Consultation, p. 9; [✂].

[590] This addresses feedback we received from one stakeholder who recommended the guidance should emphasise the need for continuous updates to proactive technologies, to ensure providers keep pace with evolving threats like new CSAM formats. A member of the Philippine Survivor Network response to the June 2025 Consultation, pp. 11-12.

[591] This addresses feedback we received from two stakeholders who argued that it was crucial proactive technology remained up to date to mitigate techniques used to evade detection by perpetrators. Scottish Government response to the June 2025 Additional Safety Measures Consultation, pp. 2-3; [✂].

[592] This addresses feedback we received from three stakeholders who welcomed Ofcom's recommendation to review existing proactive technology, as this helps avoid risks from outdated systems, supports ongoing evaluation to prevent bias or inaccuracy, and helps ensure technologies remain effective and can adapt to evolving methods used to evade detection. Advertising Standards Authority response to the June 2025 Additional Safety Measures Consultation, p. 6; Online Safety Act Network response to the June 2025 Consultation, p. 11; The Domestic Abuse Commissioner response to the June 2025 Consultation, p. 5.

it continues to be accurate, effective and lacking in bias.[593] We have received evidence from stakeholders that evaluating the performance of proactive technologies in this way is already industry best practice.[594]

9.77 We are recommending that providers review their selection of specified technologies in alignment with the risk assessment process.[595] Several industry stakeholders disagreed with the timescales we recommended providers review their proactive technology, stating this was disproportionate.[596] One industry stakeholder recommended we refine our proposals so as not to create undue documentation requirements.[597] We have aligned the frequency of the review with risk assessments. This is so that any changes to the specified technologies providers select are informed by any changes to risks occurring on the service. We consider this is a benefit to user safety as providers can make changes to their proactive technology in line with how image-based CSAM is manifesting as identified in their risk assessment to ensure it remains effective at detecting this content.

9.78 One industry stakeholder recommended giving providers discretion to decide if a significant change justified reassessing proactive technology, rather than requiring both annual and change-based reviews.[598] Where there is a change to any aspect of the service's design or operation which would amount to a significant change, this triggers a risk assessment.[599] We consider that this risk assessment may identify changes to the way image-based CSAM manifests on the service. Therefore, we consider it is important providers review their selection of specified technologies.

---

[593] Three stakeholders noted the importance of providers regularly reviewing their proactive technology to ensure they deploy the most effective tools to tackle harm. [✂]; NSPCC response to the June 2025 Consultation, pp. 25-26; Online Safety Network response to the June 2025 Consultation, p. 11.

[594] We have received evidence from stakeholders that evaluating the performance of proactive technology in this way is already best practice. Responses to our January 2025 formal information request indicate that many providers have either internal risk intelligence teams and/or rely on third party providers to monitor trends and behaviours that may impact the effectiveness of their proactive technology over time. Research also found that third-party providers of proactive technology stressed that improving the performance of a model is an on-going process, and that some providers will rely on active learning cycles, scheduled or ad-hoc updates to ensure their proactive technology is meeting performance targets – including to address new or emerging threats. See: Winder.ai on behalf of Ofcom, 2023. Automated Content Classification (ACC) Systems, p. 21; Evidently.ai, 2025. What is concept drift in ML, and how to detect and address it. In response to our June 2025 Consultation, Twitch recognised the importance of proactive technology in addressing illegal and harmful content, and that it continually assessing proactive technology systems in order to safeguard users from newly emerging risks. It also outlined their commitment and practices in continuously monitoring and evaluating their safety technology including technology already deployed as well as new technology available in the market, Twitch response to the June 2025 Consultation, p. 18.

[595] For when risk assessment(s) should be reviewed and updated: Ofcom, 2024, Risk Assessment Guidance and Risk Profiles.

[596] Google LLC response to the June 2025 Consultation, p. 16; Meta response to the June 2025 Consultation, p. 9; [✂]; [✂].

[597] Microsoft response to the June 2025 Consultation, p. 3.

[598] Google LLC response to the June 2025 Consultation, p. 16.

[599] Section 26(4) of the Act.

Benefits and effectiveness of the proactive technology criteria and draft guidance

9.79 We consider that in meeting the proactive technology criteria the specified technologies (or combinations of specified technologies) will be accurate, effective and lacking in bias for their service. In developing our proactive technology criteria:

- We have drawn on evidence from relevant national and international frameworks on responsible AI and automated content moderation.[600]
- We have considered common challenges in the design and deployment of detection technologies (such as accuracy, fairness, oversight, and adaptability) and translated these into a practical, outcomes-focused set of criteria.[601]
- While the proactive technology criteria were not adopted from a single external standard, they reflect widely recognised concerns and are consistent with guidance and recommendations published by academic, regulatory, and industry bodies.[602] [603]

9.80 We received some feedback that we should recommend certain thresholds for accuracy and effectiveness when configuring and deploying proactive technology.[604] We have not included this recommendation for the same reasons that we have not taken the approach of using metrics to assess classes/subclasses of technology for the purpose of detecting unknown image-based CSAM. For our full reasoning, see the section entitled 'Our approach to technical assessments' in annex 4.1: Proactive technology: CSAM proactive technology technical annex.

9.81 We have made a change to how the proactive technology criteria are set out in the Code for this measure. We have moved the proactive technology criteria from the 'Key definitions' section into the 'Recommendation' section, as the criteria set out substantive steps for implementing the measure. We consider setting out the proactive technology criteria in more detail, as steps to follow in the recommendation will make it easier for providers to follow and apply the criteria in practice. This is a structural change designed to improve clarity for providers and does not alter our obligations or expected practice.

9.82 We consider that the proactive technology criteria are necessary to ensure that specified technologies that are selected and used are accurate, effective and lacking in bias. This will

---

[600] NIST, 2023. NIST Risk Management Framework Aims to Improve Trustworthiness of Artificial Intelligence; Department for Science, Innovation and Technology, 2024. A pro-innovation approach to AI regulation; OECD, 2023. AI principles; European Union, 2024. Regulation - EU - 2024/1689 - EN - EUR-Lex.

[601] Two stakeholders argued the proactive technology criteria appeared to be mainly designed for reviewing AI tools. We consider the criteria are expressed in broad terms and provide practical examples of the steps service providers can take to determine whether their proactive technology is accurate, effective and lacks bias. The criteria are designed to apply across all of the classes of proactive technology we recommend. Internet Watch Foundation response to the June 2025 Consultation, p. 17; Online Safety Act Network response to the June 2025 Consultation, pp. 19-20.

[602] NIST, NIST Risk Management Framework Aims to Improve Trustworthiness of Artificial Intelligence; Department for Science, Innovation and Technology, A pro-innovation approach to AI regulation; OECD, AI principles; European Union, Regulation - EU - 2024/1689 - EN - EUR-Lex.

[603] One stakeholder provided support for the criteria. Snap Inc response to the June 2025 Consultation, pp. 5-6.

[604] A member of the Philippine Survivor Network response to the June 2025 Additional Safety Measures Consultation, pp. 11-12; Open Rights Group response to the June 2025 Additional Safety Measures Consultation, pp. 16-17; UK Finance response to the June 2025 Additional Safety Measures Consultation, p. 9; [✂].

help ensure that proactive technologies deliver their intended benefits for users. We set out our reasoning for how each criterion meets this threshold in more detail below.

High-quality data

9.83 The use of high-quality data is crucial to ensure that datasets reflect the diversity of inputs the relevant technology will encounter. If it is not, proactive technology may fail to effectively detect harmful content or behaviour; or produce outputs which have discriminatory or negative impacts for different users.

9.84 Providers should assess whether the datasets used (for training, testing or operation) are sufficiently high-quality and appropriate to the harm and service context. Providers should identify available mitigations which could be applied to address gaps or limitations.

9.85 Several stakeholders argued input from trusted flaggers and law enforcement should be used to improve proactive technologies instead of additional human moderation.[605]

9.86 We agree that trusted flagger data can sometimes be useful for training proactive technology, and we have amended our guidance related to 'High-quality data' to reflect this. We now say in the guidance that: "Providers should use data from trusted flaggers to train their relevant technology where appropriate."

9.87 We have also made a clarificatory change to this criterion to make it clear that a provider will have conducted the risk assessment with respect to a service it operates.

9.88 We consider that a reasonably high level of expertise in data testing and training is required for providers to meet this criterion. Depending on the context, providers' own in-house expertise may be sufficient for this, and both third party review of data and in-house review of data can ensure quality assured datasets.[606]

9.89 For CSAM, we have explicitly stated in the Draft Proactive Technology Guidance that providers must comply with criminal law. If service providers choose to consult with third party providers, they must comply with laws related to the handling of that kind of data.[607]

Addressing biases

9.90 Bias introduced during any stage of the proactive technology's lifecycle can lead to discriminatory or unequal treatment of different user groups, types of content or other inputs if not addressed.[608] Therefore, addressing these potential biases is an important factor to consider when configuring and deploying relevant technology to ensure users are protected.

9.91 Providers should understand how bias has been or will be identified and addressed during initial assessment, testing and configuration, development, deployment and ongoing

---

[605] Advertising Standards Authority response to the June 2025 Consultation, p.6; Financial Conduct Authority response to the June 2025 Consultation, p. 2, 7-8, [✂]; [✂]; UK Finance response to the June 2025 Consultation, p. 9.

[606] We consider that this addresses Internet Watch Foundation's argument that we recommend independent review of in-house datasets, particularly to ensure the integrity and secure handling of CSAM data. IWF response to the June 2025 Consultation, pp. 18-19.

[607] IWF response to the June 2025 Consultation, pp. 18-19.

[608] Glitch suggested that proactive technologies were often wrongly presented as neutral and inevitable solutions, but in reality were trained on biased datasets, used in opaque systems, and judged by narrow success metrics that did not reflect survivors' experiences. Glitch response to the June 2025 Consultation, pp. 3-5.

monitoring. Providers should consider how decisions made across the lifecycle of the relevant technology, such as in data selection, data labelling, rule design, or threshold setting (as relevant), may contribute to bias or unintended outcomes and identify steps to monitor and mitigate any disparities.

9.92 To ensure a provider addresses bias throughout the lifecycle of its proactive technology, we have removed the reference to the 'design and development process' from the 'Addressing biases' criterion. We have retained the stipulation that biases should be identified and addressed 'throughout the relevant technology's lifecycle'. We consider this is an important clarification to ensure the rights of users are safeguarded.

Evaluating performance

9.93 Proactive technology used for detection of relevant content involves making trade-offs between false positives and false negatives. Understanding and managing those trade-offs is essential to ensuring the relevant technology performs proportionately, balancing the risk of over-removal of legitimate content, and therefore the impact on a user's freedom of expression, with failure to effectively detect content that may cause harm to a user or to someone else (such as victims or survivors depicted in the content).[609]

9.94 This criterion aims to mitigate the risk of false positives by recommending that providers strike an appropriate balance between precision and recall in configuring their technology. We have clarified that an appropriate balance of precision and recall is a balance set with a view to minimising the incidence and impact of moderation errors, including both failures to identify image-based CSAM content and incorrect action against content. We consider this will ensure consistency in what providers decide is an appropriate balance between precision and recall on their service.[610] We consider that this consistency helps address concerns raised by stakeholders that users may not be able to sufficiently foresee the outcome of this measure as it varies from service to service (see section entitled 'Summary of stakeholder feedback' for more detail).

9.95 Service providers should take into account certain factors in striking this balance, including the service's risk of image-based CSAM, the proportion of detected content that is a false positive, the effectiveness of the systems and/or processes used to identify false positives and the importance of minimising the reporting of false positives to the National Crime Agency (NCA) or a foreign agency. These factors aim to help ensure the accuracy and effectiveness of the technology and should therefore limit the risk of detection of content that is not image-based CSAM.

---

[609] IWF said it supported using 'precision and recall' to assess AI detection tools and highlighted that false positives are a persistent issue, especially for CSAM detection due to low prevalence on most platforms. IWF response to the June 2025 Consultation, pp. 21-23.

[610] We consider this addresses Glitch's feedback that the measure's requirement for platforms to self-assess proactive technology performance lacked clear guidance on how to calculate and address performance metrics so that its performance strikes an appropriate balance between precision and recall. It argued this was a major weakness, as providers had commercial incentives to overstate accuracy and under-report harm, citing civil society research showing inflated metrics and discontinued moderation tools inconvenient to them; it concluded that self-assessment without external checks amounted to self-regulation. Glitch response to the June 2025 Consultation, p. 5.

Safeguards against misuse and exploitation

9.96 Proactive technology for content detection is susceptible to adversarial exploitation and misuse, potentially leading to compromised reliability. In order to protect users, it is important that providers ensure safeguards are in place to avoid this.

9.97 Safeguards such as input validation,[611] evasion detection[612] and configuration constraints[613] help ensure relevant technology functions as intended and maintains performance. Safeguards such as access restrictions[614] and mechanisms to ensure system and data integrity are also required to address security risks and support reliable operation throughout the lifecycle of the relevant technology.

Contextual testing and evaluation

9.98 Pre-deployment evaluation in real-world or closely simulated conditions is essential to verify the relevant technology's ability to operate effectively in the context in which it will be deployed. By testing in this way, this will ensure the technology is effective at protecting users from image-based CSAM while minimising the wrongful removal of lawful content, protecting users' freedom of expression rights. This includes its ability to operate at scale, handle relevant types of target illegal content and respond to actual usage patterns.

9.99 To ensure providers are applying this criterion effectively, we have made a minor clarificatory change stating that evaluation in this context includes "testing for scalability and handling of different media (where relevant)" and that providers should assess "whether the relevant technology's performance could be improved by layering with complementary approaches or (in the case of existing deployments) by updating to a more current version." While this change is not substantiative, we have made this change to ensure providers understand what it is they are evaluating and assessing for.

---

[611] The technical processes and controls used to verify that data or content submitted to a content moderation system conforms to predefined requirements before being processed by the system. These controls seek to identify and appropriately handle invalid, malformed, manipulated, or potentially malicious inputs that could compromise the system's functionality, interfere with its ability to detect harmful content, or enable users to circumvent moderation safeguards.

[612] The technical measures used to identify and respond to attempts to circumvent or undermine a content moderation system's ability to detect prohibited or harmful content. These measures seek to identify adversarial manipulation of inputs, suspicious patterns of activity, or other techniques intended to impact the effectiveness of the system.

[613] Technical and administrative constraints governing how a content moderation system may be configured, modified, or operated. These constraints are designed in a threat-aware way which considers the specific harm manifestation and deployment context to ensure the effectiveness, security, or compliance of the content moderation system with established safety requirements.

[614] "Access restrictions" refers to limitations placed on interaction with system operations or data, based on the roles or responsibilities assigned to any entity (including individuals, devices, or systems such as software components, applications, and automated processes), whether internal and external. This aligns with principles on controlling access, as outlined in the UK Cyber Assessment Framework, and supported by international standards such as ISO/IEC27001 and NIST SP-800-53.

Maintenance and ongoing monitoring

9.100 Ongoing monitoring and adjustment to the proactive technology are essential to maintain the effectiveness of the system over time and to prevent degradation of performance as content evolves, new manifestations of harm emerges and/or user behaviour changes.[615]

9.101 We have amended the criterion to place a greater emphasis on the relevance of quality assurance within the 'maintenance and ongoing monitoring', 'human review' and 'incorporating feedback' criterion. We consider quality assurance plays an important role in evaluating how proactive technology systems perform in practice, identifying any systemic issues and correcting identified errors, and is a key safeguard to protecting users from the wrongful takedown of content. We set out further reasoning in relation to quality assurance within the section entitled 'Steps to take on the reversal of content decisions'.

Human review

9.102 Human oversight is an essential and well-recognised safeguard in the responsible use of automated systems, to assist in ensuring the outputs are accurate, fair and contextually appropriate.[616] Human review supports accountability, protects user rights and enables corrective action where detection errors or edge cases arise. When technological outputs are explainable, meaning that it is possible to understand how and why they were generated, reviewers are better able to assess and make informed decisions.

9.103 We have amended the 'Human review' criterion to emphasise that a provider should abide by its policies and follow its processes for human review to ensure clarity around the circumstances in which human review should be used and how providers should deliver. As such, we have removed the phrase "action is taken in accordance with their policy"' and now refer to both "policies and processes".

9.104 Several stakeholders raised feedback related to the circumstances in which human review should be used, the risks associated with fully automated moderation and the impact of incorrect decisions and the need for clear and foreseeable safeguards.[617] In response to this feedback, we have clarified the factors providers should have regard to when determining an appropriate proportion for human review (both for moderation and quality assurance purposes). We outline stakeholder feedback and our reasoning for these amended factors in more detail in annex 4.3: Human review for proactive technology.

9.105 We have made a change within our human review criterion which sets out that human review serves two purposes, namely as part of moderation and for quality assurance, and when human review should be conducted within the lifecycle of the proactive technology. We set out our reasoning for this within annex 4.3: Human review for proactive technology.

9.106 We have made a change within our human review criterion and set out in our Draft Proactive Technology Guidance that, for quality assurance purposes, we consider a proper approach requires human review of detected, as well as undetected, content. We consider this will provide quality assurance as to risks of false positives and false negatives. This human review may include content that was not detected by the relevant technology but

---

[615] One provider explained how it updates its systems to reflect emerging harms. OnlyFans response to the June 2025 Additional Safety Measures Consultation, (email correspondence).

[616] One stakeholder agreed with the use of proactive technology but considered human moderation was a necessary part of the process. Scottish Government response to the June 2025 Consultation, p. 2.

[617] For more detail of this feedback, see annex 4.3: Human review for proactive technology.

has been identified as suspected target content through other means such as user reporting, complaints, community moderation, stratified sampling or other means.

Incorporating feedback

9.107 Proactive technology must adapt to remain effective and relevant in dynamic environments. Incorporating diverse and up-to-date datasets and structured feedback supports continual improvement, helps address emerging harms and/or changes to user behaviour and reduces the risk of performance degradation or bias over time. This will create more accurate and effective proactive technologies, thereby protecting users from illegal content.

9.108 It is also important that providers respond to performance issues. Therefore, we have made a clarificatory change to the 'incorporating feedback' criterion to make reference to providers identifying, correcting and responding to errors and performance issues. We consider this is an important change to protect users' rights and is intended to make it clear to providers what they should do once they have identified a false positive or a false negative.

9.109 As such, we consider that the criteria, when taken as a whole, are proportionate, apply across the specified technologies we recommend, and are adaptable to a range of service types.

Impact and costs

9.110 This section assesses the direct and indirect impacts on service providers in scope of the measure. We set out the costs of implementing proactive technology for the purpose of detecting unknown image-based CSAM using application-specific machine learning and text matching.[618]

9.111 While most elements of the cost analysis are broadly applicable to various service providers, we outline scenarios where costs may differ. We consider our assessment to be broad enough to capture an adequate range of cost scenarios in practice, while recognising that the actual implementation of the measure may vary greatly in practice in different contexts and acknowledging that it is not possible to precisely assess every real-world scenario.[619]

9.112 In our June 2025 Consultation, we estimated the cost of implementing content identification technology. We assessed costs for separate scenarios of third-party sourcing and in-house development of proactive technology, while also noting that some providers may use a combination of these approaches.

9.113 We received stakeholder responses which stated that our costs were generally underestimated and did not represent the true, ongoing and complex nature of the cost of

[618] See table 9.3 in the section entitled 'Using proactive technology' for the full list of specified technologies.
[619] We note that existing content moderation measures in the Illegal Content and Protection of Children Codes already set more general expectations for providers to have a content moderation function, to set certain policies and to resource the function appropriately. This proposed measure sets more specific expectations with regard to the use of proactive technology to address specific harms. We assess the incremental impacts of the proposed measure, over and above the existing content moderation measures assumed to already be in place.

the proactive technology measures.[620] The stakeholder responses did not identify any cost items that were overlooked in the cost assessment, nor did they provide different ranges for our estimates. Therefore, our evaluation of the costs of application-specific machine learning technology remains largely unchanged from the assessment we made of the costs of deploying content identification technology set out in our June 2025 Consultation.

9.114 However, we have made the following five updates to our consultation cost assessment:

i) We have added quantified estimates of the costs of building application-specific machine learning technology in-house, in response to stakeholder feedback.
ii) We have separated out the costs of the different specified technologies we are now recommending in this measure. We present the costs of application-specific machine learning technology and text matching technology separately. In our June 2025 Consultation, we presented one assessment designed to capture a range of proactive technologies. We have maintained this as our estimate for application-specific machine learning but have created new estimates for text matching.
iii) As explained in the section entitled 'Summary of the measure', we are now recommending specified technologies providers should use for the purpose of detecting unknown image-based CSAM. We are recommending in selecting and throughout the lifecycle of the relevant technology, the provider should ensure it meets the proactive criteria. Therefore, the nature of costs related to recording and reviewing has changed, and we have updated our presentation of these costs accordingly. We also have acknowledged the ongoing nature of some of these costs in response to stakeholder feedback.
iv) In response to stakeholder feedback about the human moderation cost stemming from the cost of false positives of relevant technology to providers, we have provided further clarity on what providers should do to achieve appropriate balance between precision and recall. We have also set out the cost implications of human review for quality assurance.
v) We have also updated our quantified estimates from the June 2025 Consultation in line with the most recent Office for National Statistics (ONS) salary data.[621]

Direct costs for service providers

9.115 The cost of implementing proactive technology depends on several factors, including the specified technologies deployed, the features of the service, and the volume and type (such as text, image, audio, or video) of content on the service to be scanned.

9.116 In our assessment below we set out the following costs:

a) The costs associated with recording and reviewing.
b) The cost of procuring and deploying proactive technology sourced from a third-party provider.
c) Our analysis of costs for a service that chooses to develop proactive technology in-house for application-specific machine learning or text matching.
d) Our assessment of costs for services with existing technology.
e) Cost impacts of human moderation.

---

[620] BILETA response to the June 2025 Consultation, p. 11; Computer and Communications Industry Association's response to the June 2025 Consultation, pp. 5-7; [✂]; GitHub response to the June 2025 Consultation, pp. 3-4; Google LLC response to the June 2025 Consultation, p. 16; [✂]; Microsoft response to the June 2025 Consultation, p. 5; [✂]; Pinterest response to the June 2025 Consultation, p. 3; [✂]; [✂].
[621] See annex 5: Further detail on economic assumptions and analysis.

f) Wider impact on the market.

9.117 We note that the distinction between 'third-party' and 'in-house' technology is a simplified approach for the purpose of our assessment. We expect that, in practice, some providers may use a combination of these approaches.

9.118 Our analysis shows that building proactive technology in-house would typically require larger investments and resources and for application-specific machine learning technologies that may only be realistic for larger services. Therefore, our assessment of the cost of implementing the proposed measure using third-party software is particularly important in our analysis of whether the measure is proportionate for smaller services.

The cost of recording and reviewing

9.119 As set out in the section titled 'Recording and reviewing', service providers should select one (or a combination) of the specified technologies to use, and decide whether to develop proactive technology in-house, source it from a third party, or use a combination of both approaches. Services are expected to incur costs in making these decisions and recording the approach taken. Where technology is sourced from a third party, there will be costs involved in identifying and selecting an appropriate supplier for the technology and deploying it, including integrating it into their service. Where technology is to be developed in house, there will be costs in planning its implementation. No matter how the technology is sourced, providers should record which technology has been selected and the evidence showing that it meets our criteria.

9.120 Across both approaches, we estimate this initial step could take one to two software engineers working alongside professionals from other occupations (such as analysts, product managers and lawyers) one to two months.[622] We estimate this labour cost to be £11,000 to £86,000.[623] Services are more likely to incur costs towards the upper end of this range if they are larger, more complex, or have multiple types of media.

9.121 Providers may also incur costs in reviewing the information recorded every 12 months, or earlier where a significant change to the service design or operation triggers a risk assessment as detailed in the section titled 'Recording and reviewing'. This would include considering whether the selected approach, supplier, classes of technology, implementation plan and evidence of compliance remain appropriate; identifying whether deployment has become technically feasible; and recording any changes made following the review. Providers may also integrate these reviews into existing processes, including annual risk assessments, and may carry them out on an ongoing basis if preferred. We have not quantified this cost as it will significantly depend on the service and their approach.

Cost of procuring and using proactive technology sourced from a third-party provider

9.122 When sourcing technology from a third-party provider, in addition to the initial costs identified in the section entitled 'The cost of recording and reviewing', providers may incur additional internal costs associated with procuring the relevant technology (for example,

---

[622] In our June 2025 Consultation, we used "software engineers - 2134" salary in ONS data for the professional occupations. Having considered further, we think that "professional occupations - 2" salary is more appropriate and reflects various other occupations might be involved in this process and is consistent with our approach to costing other measures. Therefore, updating the relevant occupation and its associated salary has slightly decreased our estimate from consultation. More details on the salary codes can be found in the annex 5: Further detail on economic assumptions and analysis.

[623] See annex 5 for further detail on economic assumptions and analysis.

making contractual arrangements and undertaking governance or budget processes), and deploying it. For smaller, less complex services, we expect such costs to be relatively low and to take less time. However, these costs may be substantial for larger services with more complex systems and technology stacks.

9.123 The technology used, and its cost may vary greatly depending on the context of a specific service. Costs are generally likely to be lower for some media types (such as text) and higher for media such as video or livestreaming content, which could exceed the indicative estimates in table 9.4 in some cases.[624] Costs may be higher where a service has multiple media types (such as text, images, video, and audio), and where the overall complexity of the service is greater (for example, where there are a larger number of different functionalities on which illegal content may be generated, shared, or encountered).

9.124 External fees paid to the third-party technology provider are likely to account for most of the cost. Fees tend to be based on the volume of content to be scanned. Additional fees, such as technical support fees, may apply in some cases. We acknowledge some extra costs that service providers might incur related to the extent of internal testing and configuration required, and any technical support or contractual fees associated with deployment.

9.125 Below, we estimate the costs of third-party application-specific machine learning and text matching based on the volume of content.

Application-specific machine learning cost estimates

9.126 Based on available evidence about third-party software within the specified technology, application-specific machine learning that has relevant capabilities for image or text detection, we assume a unit cost of £0.0005 to £0.005 for scanning one unit of content (for example, one image or 1,000 characters of text). Table 9.4 shows the illustrative annual costs for some hypothetical services to scan new content posted. For example, if a service has 700,000 users – of which 20%[625] post one new piece of content every month on average – the annual cost of scanning new content could range from £900 to £9,000.[626] On a hypothetical larger service with 7,000,000 users – of which 60% post content once a month on average – the annual cost could be £26,000 to £260,000.[627]

---

[624] For example, in the case of video content, it may be analysed by proactive technology by scanning images from multiple frames of the video, each one constituting a 'piece of content' in the sense of the indicative cost estimates of table 9.4. Therefore, costs may be significantly higher for such content, particularly where videos are long. At the same time, the measure is not prescriptive, and providers have the flexibility to explore cost effective implementation strategies, for instance in deciding any frame rate that is used for this purpose, provided that the criteria are met.

[625] We acknowledge that the volume of content posted on services will vary and have used some indicative percentage assumptions to capture plausible volumes of content posted by users on average.

[626] We expect the volume of content on a service to generally correlate with the number of monthly users. However, this relationship is unlikely to be exact and may vary significantly. On services where user engagement and posting frequency are particularly high, the volume of content is likely to be disproportionately greater. This trend would tend to drive up costs, however, this can be offset by two factors. Firstly, third-party providers often offer reduced pricing when higher volumes of content are scanned. Therefore, they would be eligible for volume discounts. Secondly, higher user engagement often correlates with greater revenue, which provides these services with a stronger ability to manage and absorb the associated expenses.

[627] We acknowledge the variability of these calculations across different services and have therefore considered different assumptions for illustrative purposes. Our aim is to capture the majority of real-world scenarios; however, we recognise that costs may still fall outside our projected ranges in certain cases,

9.127 The cost associated with scanning pre-existing content will depend on the volume of such content and could be significantly greater. For example, if a hypothetical service with a user base of 700,000 has five years' worth of pre-existing content available to UK users which has been generated at a constant rate, the cost of scanning pre-existing content could range from £4,300 to £50,000 for scanning all pre-existing text or image-based content on the service.[628]

Table 9.4: Estimated cost of scanning content using application-specific third-party machine learning technology[629]

Hypothetical service	Number of users	Percentage of users who post content once a month	Volume of new content (annual)	Estimated annual cost of scanning new content		Estimated one-time cost of scanning 5-years' worth historic content	
				Lower bound (£0.0005 per unit)	Upper bound (£0.005 per unit)	Lower bound (£0.0005 per unit)	Upper bound (£0.005 per unit)
Service A	700,000	20%	1,700,000	£900	£9,000	£4,300	£50,000
Service B	700,000	40%	3,400,000	£1,700	£20,000	£8,400	£90,000
Service C	7,000,000	40%	33,600,000	£17,000	£170,000	£84,000	£840,000
Service D	7,000,000	60%	50,400,000	£26,000	£260,000	£126,000	£1,260,000

Application specific text matching cost estimates

9.128 A service may choose to employ text matching using a third-party provider. We estimate that the costs for scanning one unit of text (up to 1,000 characters) would range between £0.0002 to £0.002, as set out in table 9.5. These costs are typically lower relative to those associated with scanning other content types such as image or video content.

9.129 The table shows the illustrative annual costs for some hypothetical services to scan new content. For example, if a service has 700,000 users, of which 20% post one new piece of content every month on average the annual cost of scanning new content could range from £400 to £4000. On a hypothetical larger service with 7,000,000 users – of which 60% post content once a month on average – the annual cost could be £11,000 to £110,000.

9.130 Costs for scanning pre-existing content will depend on the volume of such content. For example, if a hypothetical service with a user base of 700,000 has five years' worth of pre-existing content available to UK users which has been generated at a constant rate, the cost of scanning pre-existing text content could range from £1,700 to £20,000 for scanning all pre-existing content on the service.

9.131 These estimates capture the direct cost of scanning text for relevant text matches. Where a match points to a related content, services may incur additional human review, engineering

particularly for services with disproportionately high engagement or new content volume relative to their user base.

[628] We recognise that platforms may regularly archive or delete old content, reducing the cost of scanning existing data compared to stated estimates here.

[629] Figures are illustrative and rounded where appropriate.

and scanning costs to identify and assess that content. These costs will vary depending on how the service implements the text matching technology.

Table 9.5: Estimated cost of scanning content using application-specific third-party text matching technology[630]

Hypothetical service	Number of users	Percentage of users who post content once a month	Volume of new content (annual)	Estimated annual cost of scanning new content		Estimated one-time cost of scanning 5-years' worth historic content	
				Lower bound (£0.0002 per unit)	Upper bound (£0.002 per unit)	Lower bound (£0.0002 per unit)	Upper bound (£0.002 per unit)
Service A	700,000	20%	1,700,000	£400	£4,000	£1,700	£20,000
Service B	700,000	40%	3,400,000	£700	£7,000	£3,400	£40,000
Service C	7,000,000	40%	33,600,000	£6,800	£70,000	£34,000	£340,000
Service D	7,000,000	60%	50,400,000	£11,000	£110,000	£51,000	£510,000

Costs of developing and deploying proactive technology in-house

9.132 The costs of developing proactive technologies in-house are likely to vary greatly depending on the service type and will involve significant upfront costs and some ongoing costs. Below, we cost application-specific machine learning and text matching separately.

Application-specific machine learning technologies

9.133 Our assessment of the costs of developing application-specific machine learning technology in-house remains consistent with the assumptions used in our June 2025 Consultation for developing a general proactive technology, but we have now included quantified estimates of the associated upfront labour cost.

9.134 Upfront costs would include labour costs associated with designing and developing the technology; conducting thorough testing to ensure accuracy, effectiveness, and lack of bias; integrating the technology into existing systems and processes; and updating or adjusting current processes and policies. Indicatively, we estimated that undertaking the initial set-up of an application-specific machine learning based proactive technology tool could take 6 to 24 months of full-time work by a team of 4 to 6 software engineers, plus equivalent time from other professionals (such as project managers, data scientists, and lawyers). Using our standard salary assumptions, the upfront staff cost could range from £258,000 to £3,000,000.[631] [632]

---

[630] Figures are illustrative and rounded where appropriate.

[631] [✂].

[632] In our June 2025 consultation, we used "software engineers - 2134" salary in ONS data for the professional occupations. Having considered further, we have determined that "professional occupations - 2" salary is more appropriate and reflects various other occupations might be involved in this process and is consistent with our approach to costing other measures. Therefore, updating the relevant occupation and its associated salary has

9.135 There would also be non-labour costs associated with other necessary inputs and resources such as hardware and software.[633]

9.136 A provider may need to curate suitable training and evaluation data to adapt its model. As described in annex 4.1: Proactive technology: CSAM proactive technology technical annex, training and evaluation datasets present a particular challenge for unknown image-based CSAM. Providers are not expected to purchase or hold unlawful datasets containing CSAM. Providers may instead rely on testing undertaken by specialist organisations with the necessary legal permissions. Organisations including the National Centre for Missing and Exploited Children (NCMEC) and specialist safety providers have supported access to relevant datasets, training resources or governance arrangements under tightly controlled conditions.

9.137 Alternatively, providers can purchase legal proxy data for training their models.[634] Purchasing training data may be costly where a service does not have relevant training data available. This cost is likely to vary greatly depending on circumstances (for example, the type of media and the type of content). Indicatively, a large high-quality dataset could cost tens of thousands of pounds, or more. For example, the cost of acquiring a high-quality training dataset of 100,000 images could range between £100,000 and £300,000 assuming a price of £1 to £3 per image.[635]

9.138 Once the technology has been deployed, ongoing costs would involve regular maintenance and support. This includes system updates and bug fixes, as well as technology-related expenses like server management. Applying our standard cost assumptions, annual maintenance costs could range from £65,000 to £775,000 if they were equal to 25% of the initial set-up costs.[636] Ongoing infrastructure costs could also be high, particularly for large and complex services. For example, a large technology company stated that its ongoing machine costs for training machine learning classifiers (for example, compute) and data storage reaches tens of millions of dollars per year.[637]

Text matching

9.139 We estimate that the in-house set-up of a text matching tool would require between 2 to 6 months FTE work by a team of 2 to 4 software engineers plus equivalent time from other professionals (such as project managers, data scientists, legal professionals and harm

---

slightly decreased our estimate from consultation. More details on the salary codes and assumptions can be found in annex 5: Further detail on economic assumptions and analysis.

[633] We do not quantify these precisely as they are likely to vary significantly based on service providers' needs, but they are expected to be comparable to labour costs in many cases.

[634] See annex 4.1: Proactive technology: CSAM proactive technology technical annex for a description of this approach to training application-specific machine learning technologies.

[635] According to the study by Dimensional Research, 2019, Artificial Intelligence and Machine Learning Projects Are Obstructed by Data Issues, on average ML projects need around 100,000 data samples to perform well. However, this number might vary depend on the services' specific needs and structure.

[636] In our July 2025 consultation, we used "software engineers - 2134" salary in ONS data for the professional occupations. Having considered further, we have determined that "professional occupations - 2" salary is more appropriate and reflects various other occupations might be involved in this process. Therefore, updating the relevant occupation and its associated salary has slightly decreased our estimate from consultation. More details on the salary codes can be found in annex 5: Further detail on economic assumptions and analysis.

[637] [✂].

experts). Using our standard salary assumptions, we estimate that the upfront staff cost could range from £43,000 to £516,000.

9.140 When developing text matching in-house, a provider will require a list of words to match text against. If it decides to develop this internally, we estimate that it would require around 1 software engineer for 1 to 3 months, alongside 1 to 2 other professionals for around 3 to 6 months, including harm and policy experts, legal professionals, project managers and translators or language specialists).[638] We estimate that the associated staff costs for this would range between £20,000 and £151,000. There will also be costs to keep this list up to date. Alternatively, a service may opt to use a text list from a third-party provider. The costs of acquiring a list would depend on the provider chosen and the relevant membership or access fees. Examples of organisations that provide keyword lists to services include the Internet Watch Foundation,[639] Thorn[640] and the Tech Coalition.[641] The IWF, for example, state that their fees are calculated based on the industry sector and company size and that they range between £5,000 and £100,000 or more per year.[642]

9.141 Similar to application-specific machine learning, there would also be non-labour costs associated with other necessary inputs and resources such as hardware and software. We expect this to be lower than it is for application-specific machine learning.

9.142 Annual maintenance costs assumed 25% of build costs, with updated ONS data is £11,000 to £129,000. There would also be ongoing infrastructure cost and costs associated with updating the text list.

Services with existing technology

9.143 If a provider currently deploys one (or a combination) of the specified technologies in table 9.3, it should ensure that it meets the proactive technology criteria. The cost of making sure the relevant technology meets the proactive technology criteria will vary significantly between services and cannot be quantified precisely. We would generally expect that the costs to modify technology to meet the criteria would be lower than the costs to deploy this technology where it is not already used on a service.

9.144 If a provider's existing technology already meets all the criteria and other relevant requirements, confirming and documenting this should be a limited cost. Indicatively, this may require several days or weeks of work for a single employee in the case of a small service, or a few weeks for a larger, more complex service (for example, where there are a larger number of different functionalities on which illegal content may be generated, shared, or encountered).

9.145 For providers whose existing technology does not meet all the proactive technology criteria, further investment may be needed. The scale of these costs will depend on the extent of required modifications. For example:

a) If existing technology meets most of the criteria, but only partially fulfils certain other criteria, additional testing or targeted adjustments may be necessary, potentially requiring a few weeks of work.

---

[638] The resources required are likely to vary depending on the extent to which the service develops the list independently, the number of languages and dialects covered, and the characteristics of the service.
[639] IWF, n.d. Keywords.
[640] Thorn, 2024. Introducing the CSAM Keyword Hub: A Free Collection of Words Related to Child Safety Risks.
[641] Technology Coalition, n.d. Lantern: Advancing child safety through signal sharing.
[642] IWF, Keywords.

b) If existing technology and related systems or processes require more fundamental testing or adjustments to meet the criteria, then the costs may be significantly higher, potentially involving multiple months of work.
c) In cases where the most extensive changes are needed, including when a provider currently deploys relevant technology which is not covered in table 9.3, or concludes that meeting the criteria requires developing or sourcing entirely new technology, the costs are likely to be the highest. As an upper bound, costs in this case could reach the same level as those assessed above.

Human moderation cost implications

9.146 We outline our recommendations for human review in more detail in the section entitled 'How the measure works'. As described there, human moderation is expected to happen both as part of the content moderation review pipeline, and as part of the quality assurance process to review both detected and undetected content to assess the prevalence of false positives and false negatives.

Content moderation

9.147 The deployment of proactive technology may increase or decrease human moderation costs depending on the volume and nature of harmful content, existing moderation processes, and how the technology is implemented, making the overall cost impact uncertain and variable.

9.148 While proactive technology may enable automated handling of some content that would otherwise require human review, it can also identify additional harmful content that would previously have gone undetected. Therefore, it might increase the volume of material requiring review by human moderators and result in substantial costs.[643]

9.149 We have considered the responses related to the human moderation cost that emphasised the added costs of false positives of proactive technology to providers.[644] We consider we have made a number of amendments that address this feedback, however it is necessary to accept that there will be false positives. Firstly, we are currently applying this measure only with respect to image-based CSAM content, so the measure is limited to the deployment of the specified technologies we have considered in annex 4.1: Proactive technology: CSAM proactive technology technical annex. We consider these specified technologies, or combination of specified technologies, to be capable in principle of being accurate, effective and lacking in bias where they meet the proactive technology criteria. Therefore, we consider, where the relevant technology meets the criteria, the number of false positives will be minimised. Further, in deploying the relevant technology, we recommend providers to set the balance between precision and recall with a view to minimising the incidence and impact of moderation errors including both failures to identify image-based CSAM content and incorrect action against content. Further, where providers have a higher level of assurance in the detection outcomes produced by the relevant technology and any associated systems and processes, the appropriate proportion of human review may be lower, and therefore incur lower costs. For more detail, see our Draft Proactive Technology Guidance.

---

[643] As explained in the section entitled 'Text matching cost estimates', text matching may also increase human review cost where related contextual content needs to be identified and assessed. The scale of these costs will depend on how this technology is implemented.

[644] [✂]; Name Withheld 4 response to the June 2025 Consultation, pp. 3-4; [✂].

Quality assurance

9.150 Service providers are recommended to use human moderators to review an appropriate proportion of detected content to quality assure the relevant technology used for this measure. Service providers will incur costs for this that will depend on the volume of content on their service and the appropriate portion reviewed.

9.151 In addition, services providers are recommended to use human moderators to review a proportion of undetected content on their service as part of a quality assurance process. This step helps services to address the risk of significant rates of false negatives. Service providers will incur costs for this that will depend on the volume of content on their service and the appropriate proportion reviewed. The appropriate proportion is impacted by factors, including the level of assurance the provider has in the detection outcomes produced by the technology and any associated systems and processes and; the potential severity of harm to those depicted in or encountering image-based CSAM content. These factors are described in the section entitled 'How the measure works'.

9.152 For both types of quality assurance, this will provide information for a service that may enable them to identify improvements in their deployment of proactive technology. This may have the effect of enabling a service to increase the accuracy of its relevant technology, which may decrease the amount of content which needs to be reviewed in future quality assurance.

Summary of costs

Table 9.6: Summary of cost items associated with application-specific machine learning and text matching[645]

Cost category	Third-party		In-house	
	Application Specific machine learning	Text matching	Application-specific machine learning	Text matching
Reviewing and recording	£11,000–£86,000 across approaches. Includes deciding which technology to deploy, selecting the sourcing approach and recording the decision.			
Procurement	Additional cost may apply. Not quantified.	Additional cost may apply. Not quantified.	NA	NA
Supplier cost	£0.0005–£0.005 per unit of text up to 1,000 characters or one image.	£0.0002–£0.002 per unit of text of up to 1,000 characters.	NA	NA

[645] The distinction between third-party and in-house approaches is simplified for the purposes of this assessment. Providers may use a combination of the two approaches. The estimates shown do not represent total whole-life costs, as some components, including certain integration, infrastructure, human moderation, quality-assurance and review costs, are not separately quantified.

Cost category	Third-party		In-house	
	Application Specific machine learning	Text matching	Application-specific machine learning	Text matching
Technical support/ contractual fees	Additional cost may apply. Not quantified.	Additional cost may apply. Not quantified.	NA	NA
Initial development, and integration	NA	NA	£258,000–£3m upfront staff costs.	£43,000–£516,000 upfront staff costs.
Training and evaluation data	NA	NA	Potential additional costs associated with curating suitable training and evaluation data.	
Hardware and software	NA	NA	Additional non-labour costs for hardware and software. Not quantified.	
Keyword list	NA		NA	Where acquired externally, access/membership costs may apply. IWF fees of £5,000 to £100,000+. Where developed internally this could cost £20,000-£151,000.
Annual maintenance	NA	NA	£65,000–£775,000 annually, assuming 25% of build costs. Additional ongoing infrastructure cost may apply. Not quantified.	£11,000–£129,000 annually, assuming 25% of build costs, plus costs associated with updating keyword lists. Additional ongoing infrastructure cost may apply. Not quantified.
Human moderation	Human moderation costs may increase or decrease depending on the volume and nature of content, existing moderation processes and implementation. Not quantified. Additional human moderation costs for quality assurance will be incurred by services assessing an appropriate proportion of detected and undetected content for review. *Text matching may include additional human-review costs where related contextual content needs to be identified and assessed following a keyword hit.			
Ongoing review	Review at least every 12 months or earlier following a relevant significant change. Not quantified.			

Potential broader impacts for the market

9.153 In our June 2025 Consultation, we also considered the wider impacts of this measure. We explained that the proposed measure may increase costs for UK services, which could, if it led to smaller providers withdrawing from the UK, potentially reduce the diversity and quantity of content available. However, we consider that this measure is unlikely to prompt widespread withdrawal from the UK market.

9.154 We also highlighted that the measure is expected to boost demand for proactive technology, encouraging new entrants, greater competition, and innovation among technology providers. This analysis still holds.

Rights impact

9.155 The internet has fundamentally altered the way information is created, shared and consumed. Unlike traditional forms of communication, online services enable virtually any individual to publish content that can be accessed by a global audience. The internet also enables offenders to identify, contact and collaborate with one another across national borders, often with a significant degree of anonymity. In addition, the speed at which content can be disseminated online is unprecedented. Illegal and harmful content can be copied, shared and viewed by very large numbers of users in a matter of minutes, substantially increasing both the scale and persistence of harm.

9.156 Proactive technology can interfere with human rights, in particular the right to freedom of expression and privacy. However, it can also play an important role in protecting those rights, in that it may help to prevent the creation, dissemination and continued circulation of image-based CSAM.

9.157 Parliament has legislated for CSAM to be designated as "priority illegal content" under the Act, requiring service providers to use systems and processes designed to minimise the length of time for which it is present on a service. This reflects the very substantial public interest that exists in measures that reduce its prevalence and dissemination online. The use of proactive technology protects the rights of existing and potential future victims of CSAM not to be subjected to torture or to inhuman or degrading treatment, as required by Article 3 of the European Convention on Human Rights (ECHR).[646]

9.158 The use of proactive technology can also serve to protect the privacy rights of existing victims and survivors where CSAM content depicting them is being viewed, shared or otherwise circulated without their consent. In the section entitled 'Benefits and effectiveness', we describe the ongoing harm to victims caused by knowing that material depicting their abuse continues to circulate (or in some cases themselves viewing that material), or from being identified by persons who have viewed that material. Removing CSAM also helps to protect their personal data.

9.159 Measures that restrict the dissemination of CSAM online may also reduce levels of child sexual abuse, both by reducing contact offending associated with viewing or creating CSAM and by enabling offenders involved in contact offending to be identified.[647]

---

[646] The state has a positive obligation to protect children from abuse by private individuals (see *Z and others v United Kingdom* (2001) 34 EHRR 3).

[647] For a detailed analysis of the links between CSAM and further CSEA offences, see 'How CSAM and grooming offending overlap' in the Illegal Harms Register of Risks chapter titled 'CSEA'.

9.160 It is, of course, important to consider the significant implications the use of proactive technologies may have on the rights to freedom of expression and privacy of other people. These are considered in detail below.

Freedom of expression

9.161 Article 10 ECHR protects the right to freedom of expression. This right extends to the freedom to hold opinions, and to receive and impart information and ideas without interference by public authority. The protection afforded by Article 10 is broad and applies not only to ideas that are favourably received, but also to those that may offend, shock or disturb. Any interference with that right must therefore be carefully justified to ensure it is prescribed by law and necessary and proportionate to a legitimate aim.

9.162 In response to the measures proposed in the June 2025 Consultation, stakeholders raised a range of concerns regarding the impact of proactive technology on freedom of expression.

a) Some stakeholders argued that the proposed measure lacked sufficient clarity and certainty to meet the threshold of it being "prescribed by law" (see the section entitled 'Summary of stakeholder feedback').[648]
b) Some stakeholders contended that the measure was not necessary to achieve its stated aims.[649]
c) A number of stakeholders expressed concerns regarding the proportionality of the measure.[650] Several stakeholders argued that proactive technology may incorrectly identify lawful content as CSAM, resulting in false positive outcomes that interfere with lawful expression.[651] A stakeholder also expressed concerns that proactive technology may exhibit bias, whether arising from the training data used, the design of the models, or the way in which decisions are implemented in practice.[652] Some stakeholders were concerned that proactive technology may be used in future to suppress dissent.[653]
d) Some stakeholders argued that the safeguards proposed were insufficient to address the risks posed by incorrect decisions and systemic errors. Stakeholders proposed further safeguards, including enhanced human review processes,[654] expanded appeals processes[655] and greater transparency obligations.[656]

---

[648] Open Rights Group response to the June 2025 Consultation, p. 17, 20, 26; Professor Korff, D. response to the June 2025 Consultation, p. 4,6; Respondents to the Open Rights Group Campaign response to the June 2025 Consultation, p. 4; Smith, G. response to the June 2025 Consultation, pp. 1-9.
[649] [✂]; NetChoice response to the June 2025 Consultation, p. 2; Professor Korff, D. response to the June 2025 Consultation, p. 6.
[650] Are, C, Dr. response to the June 2025 Consultation, p. 15; Francis, S. response to the June 2025 Consultation, pp. 7-8; [✂]; Google LLC response to the June 2025 Consultation, pp. 13-14; NetChoice response to the June 2025 Consultation, p. 2; Professor Korff, D. response to the June 2025 Consultation, p. 12; [✂]; Together Declaration Campaign response to the June 2025 Consultation, pp. 4-5; [✂].
[651] BILETA response to the June 2025 Consultation, p. 8; European Digital Rights Organisation response to the June 2025 Consultation, p. 1; [✂]; Open Rights Group response to the June 2025 Consultation, p. 16.
[652] Glitch response to the June 2025 Consultation, p. 6.
[653] [✂]; Together Declaration campaign response to the June 2025 Consultation, pp. 4-5.
[654] [✂]; South West Grid for Learning response to the June 2025 Consultation, p. 7.
[655] Are, C, Dr. response to the June 2025 Consultation, p. 15; Open Rights Group response to the June 2025 Consultation, pp. 14-15.
[656] A member of the Philippine Survivor Network response to the June 2025 Consultation, p. 12; Internet Watch Foundation response to the June 2025 Consultation, p. 18; [✂].

9.163 We carefully considered these points when reaching our decision. In particular, we considered the new content moderation measure (see section entitled 'Steps to take on reversal of content decisions'), the widening of appeal mechanisms and whether to adopt additional safeguards beyond this. Where particular suggestions have not been adopted, they are addressed either in the main reasoning or below.

Is there an interference?

9.164 We recognise that the measure interferes with the Article 10 ECHR rights of both users and service providers. Given the central role that online services play in facilitating expression, we acknowledge that this interference can be significant.

9.165 To the extent that the proactive technology correctly detects image-based CSAM, interference with freedom of expression does not raise concerns because the content in question is illegal. We are mindful that some nuances arise in relation to SGII shared by children within age-appropriate and apparently consensual relationships. However, such content is illegal.

9.166 However, where proactive technology incorrectly detects lawful content as unknown image-based CSAM, it can give rise to substantial freedom of expression concerns. The extent of the interference will depend on the way that the provider has deployed proactive technology as part of a content moderation function, including whether removal of content detected by the proactive technology is automatic or subject to additional layers of review before removal. The detection of content by proactive technology gives the provider a reason to suspect the content in question may be CSAM and, in combination with the content moderation measures, may result in the provider incorrectly taking down the detected content. Incorrect identification could affect lawful expression of all kinds, including political, religious, educational, creative and personal expression. The consequences of a false positive, an interference with freedom of expression, extend beyond the removal of content. Users may also be subject to sanctions imposed by the service, restrictions on account functionality or, in some cases, referrals to law enforcement bodies. In particular, we have recommended that for users not determined to be children who post CSAM, the default sanction should be a permanent ban from the service.

9.167 There is also a risk that the measure may have a chilling effect and users may self-censor. Some individuals may choose not to post lawful content because they fear that it may be incorrectly identified as CSAM.

Prescribed by law

9.168 The measure must be sufficiently foreseeable and accessible to satisfy the requirement that any interference with Article 10 rights be prescribed by law. Accessibility requires that a rule must be published and comprehensible to affected parties, thereby allowing that person to understand what the rule entails. That is clearly the case here. The foreseeability requirement is that it must be possible to predict how a law or rule will be applied in practice. This does not mean that Ofcom must codify the answers to every possible issue which may arise.[657] As set out in the section entitled 'Benefits and effectiveness' it is not possible (and nor would it be desirable) for the regulator to specify metrics to be applied. Nor can we set out for lawful users how to avoid their content being accidentally picked up by proactive technology. The answer would be different from service to service and from

[657] *R (Catt) v ACPO* [2015] AC 1065 at paragraph 11. The courts have acknowledged the risk of excessive rigidity – see e.g. *Satakunnan Markkinapörssi Oy & Satamedia Oy v Finland* (2018) 66 EHRR 8 at paragraph 143.

time to time and in any event would teach offenders how to evade detection. Yet the Act does foresee that Ofcom should be able to recommend the use of proactive technology where appropriate.

9.169 We have specified the type of illegal content which providers should seek to detect using proactive technology and explained in annex 4.1: Proactive technology: CSAM proactive technology technical annex why we consider it to be detectable. Having had careful regard to the responses to the June 2025 Consultation, we have recommended specified technologies which should be used to detect it. We have specified the objective for which providers should strive in determining the appropriate balance of precision and recall: a balance set with a view to minimising the incidence and impact of moderation errors including both failures to identify image-based CSAM content and incorrect action against content. Users will know from the terms and conditions of the service that proactive technology is being used and how it works.[658] We will give guidance on the application of the proactive technology criteria. We set out a series of safeguards against incorrect detections (see the section entitled 'Proportionate'). We therefore consider that the measure is sufficiently foreseeable to be prescribed by law.

Necessary for a legitimate purpose

9.170 The measure pursues the legitimate aims set out in Article 10(2) ECHR of preventing and detecting crime. Detecting CSAM as priority illegal content and removing, and reporting, CSAM acts to prevent crime by deterring users from posting such content and by preventing other users from accessing it. It also acts to protect public morals, including by preventing users from inadvertently encountering CSAM. It protects children and victims of crime, the most vulnerable members of society.

9.171 The measure goes no further than is necessary to pursue those aims. As set out in the section entitled 'Benefits and effectiveness', reliance on user complaints or manual detection processes cannot address the scale of the harms, which are very severe. Proactive technology is therefore necessary. Neither proactive technology nor human review can be perfectly accurate, but human review alone cannot address these harms at the necessary scale and speed.

9.172 To address this serious harm, it is necessary to accept that there will be false positives.

Proportionate

9.173 We consider that the risk of false positives is proportionate, given the seriousness of the harm. As a preliminary point we note that the measure only applies in relation to content communicated publicly (consistent with the constraints on Ofcom's power to include 'proactive technology' measures in a Code of Practice). The measure is limited to recommending the deployment of the kinds of technology we have considered in annex 4.1: Proactive technology: CSAM proactive technology technical annex, and in deploying it we recommend providers to set the balance between precision and recall with a view to minimising the incidence and impact of moderation errors including both failures to identify image-based CSAM content and incorrect action against content.

9.174 In principle, adverse impacts on freedom of expression could arise in relation to the most highly protected forms of content, such as religious and political expression, and in relation

[658] Section 10(7) of the Act.

to kinds of content that the Act seeks to protect, such as content of democratic importance or journalistic content.

9.175 However, the measure is unlikely to have a systematic impact on these kinds of content – we see no reason to consider that such content is more likely than other content to be incorrectly detected as a false positive, or to be concerned that any particular political message would be suppressed in this way. We expect, to the extent bias occurs, notwithstanding the criteria we have designed which are intended to prevent it, it is most likely to arise from biases in training data (for example, overrepresentation of white people leading to errors in identifying children of other races). Nothing in the measure gives any grounds upon which an arm of the state could intervene to deliberately suppress criticism or debate. The measure is also unlikely to result in the disclosure of journalistic sources to law enforcement authorities, given that it concerns only images and videos and applies only to content communicated publicly.

9.176 Providers have a reputational and commercial incentive to strive for as many true positives and as few false positives as possible.

9.177 Most importantly, the measure is subject to a range of safeguards which work to secure that both the rate and the consequences of errors are mitigated to the extent reasonably possible.

9.178 The measure incorporates safeguards relating to the deployment of proactive technology in our Proactive Technology Criteria. These include:

a) Testing and evaluation of the technology, including in context,
b) Quality assurance processes which secure ongoing monitoring and maintenance
c) Feedback on the outcomes achieved, and mechanisms to incorporate it so as to maintain or improve performance over time,
d) Identification and monitoring of potential biases,
e) Human review of an appropriate proportion of detected content, having regard to the level of assurance in the detection outcomes achieved by the technology, the harm to those depicted in or encountering illegal content and the impact on users of incorrect detections,
f) Security arrangements to maintain the integrity of the detection system, and
g) Record keeping so that decisions can be reviewed and understood.

9.179 Additional safeguards exist in the form of our existing content moderation measures for many providers, including maintaining an internal content policy, performance targets, materials and training.

9.180 Users who have been affected by errors, i.e. a false positive detection of CSAM, have access to appeals processes regarding content decisions, account restrictions and sanctions.

9.181 Moreover, there are additional complaint routes relating to providers' compliance with online safety duties and rights to freedom of expression that may provide a means by which users can alert providers to dissatisfaction, although we recognise that these mechanisms may offer more limited redress than appeals processes.

9.182 Providers also have commercial incentives to minimise errors because inaccurate decisions may undermine user trust and engagement, and because of the burden of considering appeals.

9.183 The Act requires that users be made aware via providers' terms of service that proactive technology is being used on the service, and how it works.

9.184 Where errors are identified, providers should implement mechanisms to ensure that similar mistakes are not repeatedly made and that systems are continually improved. We have introduced a new measure in our content moderation measures to act as a safeguard, which ensures that errors in content moderation decisions, where they are not identified through a content appeal, are reversed and, as far as possible, providers ensure the same error does not happen again see the section entitled 'Steps to take on reversal of content decisions'.

9.185 Data protection legislation provides an additional layer of protection, including requirements relating to fairness, accuracy, transparency, and accountability.

Conclusion

9.186 CSAM causes exceptionally serious and enduring harm to children. Given the scale at which such content can be disseminated online, it would be extremely difficult to address this harm effectively without the use of technology capable of identifying it proactively and disrupting it.

9.187 We acknowledge that the measure may have significant impacts on the freedom of expression rights of some users and service providers. However, we consider that those impacts are outweighed by the importance of protecting existing and future victims of child sexual abuse.

9.188 Overall, while we acknowledge that the measure interferes with freedom of expression where lawful content is incorrectly identified as CSAM, we consider that interference to be limited in practice by the safeguards that are in place and the residual interference to be proportionate to the legitimate objective of reducing the prevalence and dissemination of CSAM.

Privacy

9.189 Article 8 ECHR protects the right to respect for private and family life, home and correspondence. The concept of privacy is broad and encompasses personal autonomy, personal identity, bodily integrity, personal relationships and the protection of personal information.[659] The right is engaged whenever individuals have a reasonable expectation of privacy. It is engaged when personal data is collected, analysed, retained, disclosed or otherwise processed in a manner that affects an individual's private sphere.[660]

9.190 Not every aspect of data protection legislation is synonymous with the right to privacy under Article 8 ECHR. Data protection obligations may, in some circumstances, extend beyond matters that directly engage Article 8. Nevertheless, data protection rights perform an important protective function and operate as a significant safeguard for privacy.[661] Compliance with data protection legislation is therefore relevant to assessing the extent of any interference with privacy rights and the proportionality of that interference.

9.191 Stakeholders raised a number of concerns regarding the impact of proactive technology on privacy rights.[662] We do not repeat concerns already addressed elsewhere in this document or in our assessment of freedom of expression in the section entitled 'Freedom of

---

[659] *R (Countryside Alliance and ors) v Attorney General & Anor* (2007) UK HL 52

[660] *Campbell v United Kingdom* (1993) 15 EHRR 137

[661] ICO response to the June 2025 Consultation, p. 12.

[662] 5Rights Foundation response to the June 2025 Consultation, p. 9; ICO response to the June 2025 Consultation p. 6, 7, 12,14; [✂]; [✂].

expression'. Instead, we focus here on additional privacy-specific issues raised during consultation, including points raised by the Information Commission's Office (ICO) and other stakeholders, to the extent that they relate to the measure we are now recommending.

9.192 Three stakeholders raised concerns about perceived overreach or surveillance implications of deploying proactive technology, or that it may constitute general monitoring.[663] Stakeholders also raised concerns that automated systems may produce biased outcomes.[664] Such concerns were raised both as a matter of privacy and as a matter of fairness.[665]

9.193 Multiple stakeholders provided feedback about the data protection implications of our proactive technology proposals.

a) The ICO said service providers should implement the measure in line with the lawfulness, accuracy, fairness, and transparency principles of data protection legislation.[666]
b) Concerns were raised about the use of proactive technology as a form of automated decision-making, within scope of Articles 22A-D UK GDPR, that would require particular safeguards in its deployment.[667]
c) Stakeholders highlighted that the deployment of machine learning systems may involve the processing of special category personal data.[668] The ICO provided feedback that decisions concerning whether content is unlawful may involve the processing of criminal offence data.[669] It argued that such processing requires particular caution because of the potentially significant consequences for affected users and service providers should be mindful of the requirements of data protection legislation, including the principles of accuracy and fairness, when taking decisions that may have a negative impact.[670]
d) Several stakeholders emphasised the importance of the principle of data minimisation when proactive technology is deployed and ensuring any proactive technology is not over-collecting data about users.[671]
e) 5Rights Foundation recommended alignment between the deployment of user profiling and behavioural identification forms of proactive technology and the protections for children under ICO's Age Appropriate Design Code.[672]
f) A number of stakeholders also raised concerns about increased security risks and risks of data breaches and cyber-attacks where proactive technology is deployed,

---

[663] Glitch response to the June 2025 Consultation, pp. 11-12; Microsoft response to the June 2025 Consultation, p. 3, [✂].
[664] [✂]; Glitch response to the June 2025 Consultation, pp. 11-12; [✂]; Open Rights Group response to the June 2025 Consultation, pp. 45-46; NSPCC response to the June 2025 Consultation, pp. 52-54.
[665] ICO response to June 2025 Consultation, p. 10.
[666] ICO response to the June 2025 Consultation, p. 8.
[667] Professor Korff, D. response to the June 2025 Consultation, pp. 9-10.
[668] Professor Korff, D. response to the June 2025 Consultation; ICO response to the June 2025 Consultation, p. 12; [✂].
[669] ICO response to the June 2025 Consultation, p. 12.
[670] [✂]; ICO response to the June 2025 Consultation, pp. 3, 9, 11-12.
[671] 5Rights Foundation response to the June 2025 Consultation, p. 9; Glitch response to the June 2025 Consultation, pp. 11-12; [✂].
[672] 5Rights Foundation response to the June 2025 Consultation, p. 9.

highlighting the importance of the sixth data protection principle.[673] Several also mentioned that service providers should be looking to use privacy-preserving and cryptographic methods of processing personal data in proactive technology.[674]

g) The ICO also recommended a number of actions to promote accountability: that all parties are clear about their respective roles and responsibilities under data protection legislation, including who is a controller and who is a processor, and that a suitable Data Protection Impact Assessment is carried out.[675]

9.194 The ICO also emphasised the importance of the Proactive Technology Criteria and its own published guidance in ensuring compliance with data protection legislation when implementing the measure.[676]

9.195 We considered all these points when reaching our decision. We considered the safeguards for privacy that we are recommending as part of this measure, as well as the ways in which we can signpost to relevant ICO guidance (noting Ofcom's remit is not to act as a data protection regulator). In particular, we considered the new content moderation measure, the widening of appeal mechanisms and whether to adopt additional safeguards beyond this.

Is there an interference?

9.196 As noted above, this measure recommends the use of one or a combination of application-specific machine learning and text matching technologies on content communicated publicly. Where service providers ensure that this automated processing is carried out in compliance with data protection law, its impact on the privacy of users acting lawfully will (in most cases) not be very big. The processing will involve the automated use of technology to process images and text to find matches. Most content will not be detected as suspected CSAM.

9.197 Where content is detected as suspected CSAM, there will be further processing. In particular, review of some detected content by human moderators may have a different, and more significant, impact on users' privacy.[677] It will trigger a statutory reporting obligation to law enforcement, possible criminal investigation and prosecution, as well as probable sanctions from the service provider which impact on the user's ability to interact with family and friends using the service. Users who did not post the content, but who interacted with it without reporting it, could also experience sanctions which significantly affect their ability to interact with family and friends using the service, as well as possible criminal investigation and prosecution. This would be an interference with their Article 8 ECHR rights but would not be an unreasonable one where it is a necessary and

---

[673] [✂]; [✂].

[674] Glitch response to the June 2025 Consultation, pp. 11-12; Internet Watch Foundation response to the June 2025 Consultation pp. 23-24; [✂].

[675] ICO response to the June 2025 Consultation, p. 7.

[676] ICO response to the June 2025 Consultation, p. 7, 9, 10, 14.

[677] Review by human moderators of content accurately detected to be CSAM represents a significant interference with the privacy rights of the victims it depicts. However, that review forms an important part of ensuring that the overall measure – which aims to protect victims, including from serious impact on their privacy from the further dissemination of CSAM – is proportionate and appropriate for service providers to take for the purposes of complying with their illegal content safety duties. We therefore consider that the intrusion into victims' privacy rights is necessary and that no less intrusive approach would be a suitable alternative.

proportionate one with the legitimate aim of preventing and detecting crime. There is no reasonable expectation of privacy in the commission of illegal acts.[678]

9.198 A different analysis applies where lawful content is incorrectly identified as CSAM. Individuals depicted in such content may reasonably regard the review, analysis or retention of their images as an unexpected and intrusive form of processing. The intrusion may be particularly acute where highly personal, intimate or sensitive content is involved.

9.199 Users who post lawful content that is detected by proactive technology will also experience an interference with their privacy rights, as the personal data processing would not have occurred without the incorrect detection.

9.200 However, the measure is limited to content communicated publicly. By choosing to publish content on a public-facing service, users necessarily reduce, to some extent, their reasonable expectation of privacy in relation to that content. While the content remains protected by Article 8, the extent of the interference must be assessed in light of the fact that the information has been made available to a potentially wide audience and in the knowledge that services moderate content in line with their terms of service and the law.[679]

9.201 The operation of proactive technology may require data to be retained for a period of time to facilitate detection, review, quality assurance, auditing or appeals processes. The retention of personal information is itself an aspect of data processing that engages Article 8 rights and the data protection principle of storage limitation.

9.202 Where the operation of a proactive technology system involves the processing of information that constitutes criminal offence data or special category data under data protection legislation, the intrusion may be particularly significant. Such categories of information attract enhanced protection under data protection law because of the potential consequences that may arise from their misuse, misinterpretation or inappropriate disclosure.

In accordance with law

9.203 We consider that any interference with privacy rights resulting from the measure is sufficiently in accordance with the law. The measure is accessible, foreseeable and subject to a framework of legal safeguards. Our reasoning concerning the freedom of expression "prescribed by law" requirement is set out in full above in the section entitled 'Prescribed by law' and applies equally in the privacy context in relation to the criterion of "in accordance with the law".

Necessary for a legitimate aim

9.204 As explained above, the measure pursues the legitimate aims as set out in Article 8(2) ECHR of preventing crime (by detecting illegal content amounting to one or more of the CSAM-related offences in Schedule 6 to the Act) and protecting the rights and freedoms of others. These are particularly weighty objectives given the serious and enduring harms associated with CSAM.

9.205 We consider that the measure goes no further than is necessary to achieve those objectives. Importantly, the measure does not prescribe the specific data inputs that providers must use. Instead, it requires providers to undertake rigorous testing and

---

[678] *Sutherland v HM Advocate (Scotland)* [2020] UKSC 32

[679] Service providers will also be subject to the data protection principle of fairness, to ensure that such interference is not unjustified.

evaluation in line with the proactive technology criteria before deploying proactive technology. This approach allows providers to identify solutions that are effective while minimising unnecessary processing of data, respecting the data protection principle of data minimisation.

9.206 Providers remain subject to the requirements of data protection legislation when implementing the measure. These requirements include obligations relating to fairness, transparency, necessity, proportionality, purpose limitation, storage limitation, and data minimisation. Those obligations help ensure that any processing remains confined to what is necessary and appropriate. Providers should refer to the guidance published by the ICO as to compliance with their obligations under UK GDPR.

Proportionate

9.207 Although the proactive detection of image-based CSAM inevitably involves a form of processing of user-generated content and personal data, we consider that the mitigation of a serious societal harm and the privacy rights of victims and survivors together outweigh the interference with users' privacy, in light of the safeguards built into the measure.

9.208 The existence of effective safeguards is central to our assessment of the measure's proportionality. Many of the safeguards discussed in the section entitled 'Freedom of expression' also reduce privacy risks. In particular, we have incorporated safeguards designed to minimise false positives, which reduce the likelihood that lawful user-generated content and associated personal data will be unnecessarily processed, retained or disclosed.

9.209 The Proactive Technology Criteria requirements relating to testing, validation, monitoring, accuracy, and human oversight provide important protections against inappropriate or excessive processing of personal data. These measures reduce the risk that users will be adversely affected by inaccurate processing or automated decisions that unfairly impact them.

9.210 Other complaints and redress mechanisms, including the ability to make complaints under data protection legislation, provide users with an opportunity to challenge decisions that interfere with privacy and bring data protection concerns to the attention of providers. Although these mechanisms may not fully remedy every privacy impact, they constitute an important accountability safeguard.

9.211 Providers are generally required to explain their content moderation practices, use of proactive technology, and measures to uphold user privacy through their terms of service, privacy notices and related documentation. As a result, users are more likely to understand the circumstances in which their content may be reviewed, analysed or processed.

9.212 Providers also have commercial incentives to avoid erroneous, excessive or intrusive processing. Loss of user trust and reputational damage may arise if providers fail to deploy technologies responsibly.

9.213 Data protection legislation provides a further layer of protection. Regulatory requirements relating to lawfulness, fairness, transparency, accountability, data minimisation, security and individual rights significantly constrain the manner in which providers may implement the measure.

9.214 We acknowledge that the privacy intrusion associated with content analysis may be significant, particularly where sensitive images or information are involved. However, the processing is generally carried out by private companies operating online services rather

than directly by the state. While this does not eliminate the interference, it is a relevant consideration when assessing its overall weight.

9.215 Users of online platforms ordinarily expect that content they choose to publish will be subject to some degree of automated and human moderation. They also expect that information provided to platforms will be processed for operational, safety and compliance purposes, as described in the terms of service and privacy notice of the relevant platform. These expectations do not remove the protection afforded by Article 8, but they are relevant to assessing the extent of the interference and the reasonableness of the expectation of privacy.

9.216 Moreover, where content has been published on a public forum, the user has voluntarily shared that content to a potentially large audience. In those circumstances, the expectation of privacy in relation to the published content is lower than it would be for private communications or privately stored material. The measure only applies to content communicated publicly, not to content communicated privately.

Conclusion

9.217 The measure is directed towards addressing exceptionally serious harms associated with the creation and dissemination of CSAM. Those harms affect the dignity, autonomy, safety and privacy of children and can continue for many years after the original abuse has occurred. We recognise that the measure involves meaningful interferences with privacy rights, including through the processing of images, the possibility of human review, the retention of information, and the potential disclosure of information to law enforcement authorities. These impacts are particularly significant where content is incorrectly identified or where sensitive personal information is processed.

9.218 Overall, while we acknowledge that the measure involves some interference with users' right to privacy, particularly where lawful content is incorrectly identified as CSAM, we consider that interference to be limited by the safeguards that apply to the design, deployment and operation of proactive technology. In light of the severity of the harms being addressed, the lack of availability of less intrusive means for mitigating the harm, the safeguards in data protection legislation, and the protections available to users through our other measures including the steps to take on reversal of content decisions, we consider the measure to be both necessary for and proportionate to the legitimate aim of reducing the prevalence and dissemination of CSAM and protecting the rights of children and other affected individuals.

Who the measure applies to

9.219 As explained in 'Our decision', we are proceeding with the measure for the purpose of detecting unknown image-based CSAM. In our June 2025 Consultation, we proposed the measure should apply to:

a) Large user-to-user services that are medium or high risk of image-based CSAM.
b) User-to-user services with more than 700,000 monthly UK users that are high risk of image-based CSAM.
c) User-to-user services that are file-storage and file-sharing services of any size that are high risk of image-based CSAM.

9.220 Having considered stakeholder responses, we have decided to retain this approach and expand on our reasoning for including each type of service in the following paragraphs.

9.221 As set out in the section entitled 'Benefits and effectiveness' we recognise that our definition of the risk of harm is slightly broader than the definition of the content we are asking providers to detect, but we do not consider that this will have any practical impact.

Large services at medium or high risk of image-based CSAM

9.222 Given the scale of content on these services, relying solely on user reporting and complaints, and human moderation is unlikely to detect image-based CSAM effectively before it reaches large numbers of UK users and causes significant harm.

9.223 While the costs of this measure are likely to be significant, applying the measure to large, medium risk services is expected to lead to a material reduction in the relevant harm and deliver significant benefits to UK users. Taking these benefits into account and bearing in mind the extreme severity of the harm CSAM causes, we consider the costs of the measure are justified and consider it to be proportionate for it to be applied to large medium and high risk services.

Services with 700,000 or more monthly UK users at high risk of image-based CSAM

9.224 Although we recognise that the implementation costs are likely to be more burdensome for smaller services within this size category, the severity and reach of the harm to users remains significant. Additionally, we consider costs to scale with the volume of content.

9.225 Given the high level of CSAM risk on these services and large volumes of content they host, relying on user reporting, complaints and human moderation alone is unlikely to be sufficient to detect and address the harm effectively. As we have explained above, the deployment of proactive technology will materially increase the amount of CSAM that is detected and removed. Considering the extreme harm CSAM causes, we consider that applying the measure to high-risk services with 700,000 or more monthly users is justified by the expected benefits to user safety.

File-storage and file-sharing services at high risk of image-based CSAM

9.226 We have not set a size threshold on file-storage and file-sharing services with high risk of image-based CSAM. Given the severity of harm CSAM causes and the prolific amounts of CSAM that even small file-sharing services can host,[680] we consider it proportionate to apply this measure to all such services.

9.227 We received some feedback from stakeholders about the privacy issues around including these services in scope.[681] As outlined in the sub-section titled 'Privacy' in the section entitled 'Rights impact', only content communicated publicly on these services is subject to the proactive technology. For further evidence and reasoning on their inclusion, see the

[680] As set out in our December 2024 Statement, many file-storage and file-sharing services allow users to store and share vast numbers of files, meaning that small numbers of users could use the service to disseminate CSAM at scale. Moreover, as discussed in the Illegal Harms Register of Risks, some services with smaller user bases offer users specific functionalities which may not be available on services with larger user bases, such as the ability to post content without a registered account. As such, perpetrators may target these kinds of smaller services to exploit such functionalities. For further information, see our December 2024 Statement in the chapter entitled 'Automated content moderation'.

[681] European Digital Rights Organisation response to the June 2025 Consultation, p. 1; Microsoft response to the June 2025 Consultation, p. 4; Open Rights Group Response to the June 2025 Consultation, pp. 18-19.

chapter entitled 'Automated content moderation' in the December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement).[682]

Our considerations on search services

9.228 Several stakeholders said that our measures should be applied to search service providers, or that we should further explore evidence to do this.[683]

9.229 We are confirming our decision proposed in our June 2025 Consultation not to apply the measures to search service providers for the same reasons we outlined then:

i) First, we have previously received feedback from stakeholders that the complexity of automated moderation of organic search content means that we cannot necessarily transfer principles developed for user-to-user services across to search services.
ii) Second, we have limited information on how moderation is applied to each 'layer' of search (crawling, indexing, serving results) that would enable us to develop appropriate criteria for the measure.
iii) Third, we do not consider our evidence base to be robust enough to recommend all search services deploy this technology at this time.

9.230 However, we agree with stakeholders that we should continue to build our evidence base for how proactive technology measures could apply to search services and may consider expanding the measures to search service providers in the future.

Conclusion

9.231 The severity of unknown image-based CSAM and its increasing prevalence online means the deployment of proactive technology is necessary to prevent UK users from encountering this harm. While user reporting and complaints systems are important in detecting illegal content and can form part of a wider safety strategy, we do not consider this alone is sufficient to mitigate the risk of image-based CSAM. The existing measures on hash matching in the Illegal Content Codes are useful. However, by their nature, they focus on the detection of previously identified CSAM. This proactive technology measure enables the detection of unknown image-based CSAM.

9.232 We recognise that recommending proactive technology poses interference to users' rights, including their freedom of expression and privacy. We have tried to mitigate this through including various safeguards in the measure. However, we recognise interferences may remain. Given the severity of harm of image-based CSAM, and a lack of availability of a less intrusive means of mitigating this harm, we consider that this measure is still necessary for, and proportionate to, the aim of reducing the prevalence and dissemination of CSAM.

9.233 Given the benefits associated with this measure, we also consider the costs impacts on service providers to be proportionate.

9.234 We therefore conclude that this measure will deliver significant benefits in reducing the risk of harm caused by image-based CSAM and will deliver protections to children and survivors of child sexual abuse. As such, we have decided to proceed with this measure as outlined in

[682] Ofcom, 2024. Illegal Harms Statement: Volume 2 Service Design and User Choice, p. 192.
[683] Digital Resilience in Education – Welsh Government response to the June 2025 Consultation, p. 7; Evangelical Alliance's response to the June 2025 Consultation, p. 4; Financial Conduct Authority response to the June 2025 Consultation, pp. 2, 7-9; [✂]; [✂].

the section entitled 'Our decision'. This measure is part of the draft Illegal Content Code of Practice for User-to-user Services: Proactive Technologies (unknown content) and is referred to as ICU PT1.

Steps to take on reversal of content decisions

How these measures work and our reasoning

9.235 In response to the June 2025 Consultation, one industry stakeholder and one civil society stakeholder highlighted the role of human review and quality assurance processes to ensure the accuracy, fairness and consistency of proactive technology outcomes.[684] Several stakeholders argued that content and/or accounts should be reinstated where proactive technology had incorrectly taken down legal content.[685]

9.236 We agree with stakeholders that quality assurance plays an important role in evaluating how proactive technology systems perform in practice, identifying any systemic issues and correcting identified errors. To be effective, quality assurance measures should also be capable of addressing any quality issues identified in the content moderation process.[686]

9.237 We also agree that, where a provider identifies that it has incorrectly taken action on content or a user, it should reverse such actions so far as possible and appropriate.[687] Therefore, in response to stakeholders' feedback we have added a measure to our content moderation and search moderation measures in the Illegal Content and Protection of Children Codes of Practice to ensure that decisions are reversed so far as possible and appropriate where the provider finds them to be false positives. We expect this to be relevant principally when a provider carries out quality assurance, but it is possible that a provider might learn of an error in other ways (for example, if notified of a court decision).

9.238 As is already the case when a provider reverses a content decision following an appeal, we consider that a provider which reverses a content decision following any other kind of review should take steps where possible and appropriate to secure that the same errors are not (unnecessarily) made again. Where there is a pattern or significant evidence of errors, providers should, where appropriate, adjust any relevant content or search moderation guidance to improve its accuracy.

---

[684] Glitch response to the June 2025 Consultation, p. 12; Online Responsibility Network response to the June 2025 Consultation, p. 3.

[685] Open Rights Group response to the June 2025 Consultation, p. 29; Respondents to the Open Rights Group Campaign response to the June 2025 Consultation, p. 2; Professor Korff, D. response to the June 2025 Consultation, p. 13.

[686] In our Draft Proactive Technology Guidance, we set out that providers should use quality assurance to: Identify the prevalence, causes and patterns of false positives and false negatives; assess whether relevant technology is appropriately calibrated, as evidenced by observed error rates, performance variation and outcomes from quality assurance; identify inconsistencies in the performance of the relevant technology across harm categories, content types or user groups; identify any novel content formats, emerging trends or indications of adversarial behaviour which the relevant technology has not been trained to detect; and assess whether detection errors indicate limitations in training datasets, configuration of the relevant technology and related systems or human review processes. For further detail, please see the Draft Proactive Technology Guidance, in the criterion entitled 'Maintenance and ongoing monitoring.'

[687] We recognise that this may not always be possible or appropriate, for example because content is ephemeral or because it breaches providers' terms of service in other ways.

9.239 Similarly, a provider should take appropriate steps to secure that the use of automated content moderation technology does not cause the same content to be actioned, or the same search content to no longer appear in search results or be given a lower priority in the overall ranking of search results, where possible and appropriate. To mitigate this, providers should consider the need to adjust their automated content moderation technology where possible and appropriate do so. We recognise adjustments will not necessarily be needed in all situations, particularly where doing so would fail to prevent the same inaccuracy or result in the technology to not work as appropriate to protect users.

9.240 As we have explained previously in our December 2024 Statement and April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement), reversing incorrect decisions (for both content and accounts) is critical to protecting users' rights to freedom of expression and data protection rights to accurate processing of their personal data.[688] It may also encourage providers to perform a more thorough review in their initial judgements in order to avoid having to process false positives.

9.241 In terms of the costs for service providers, we note that all services in scope of this measure will also be in scope of measures in our Illegal Content and Protection of Children Codes that set out that a providers should reverse actions following a successful appeal. We consider that there are likely to be overlaps in the systems and processes that providers would use for meeting those measures and what they would use to reverse actions because of this measure, limiting the additional cost burden of this measure. We also consider the costs of this measure to be limited because we expect that the content identified will primarily be from service's quality assurance activities. The cost of appropriate steps to secure that the use of automated content moderation technology does not cause the same content to be incorrectly actioned is highly dependent on the technology used and what steps would be appropriate.

9.242 Where the service in question is in scope of our measure on deployment of proactive technology to detect image-based CSAM (ICU PT1), our Draft Proactive Technology Guidance sets out how providers should deploy and configure this technology, including incorporating feedback from quality assurance. The impacts of this are assessed in the previous section titled 'How the measure works' of this chapter.

9.243 We consider that it is appropriate to recommend this measure to all user-to-user and search services.[689] [690] This aligns with the approach taken in our measures on taking appropriate action following successful appeals (ICU D10/ICS D9 and PCU D10/PCS D10). We consider this proportionate for all services, as a provider which does not have a quality assurance process will likely be unaware of content it will need to correct a decision for.

---

[688] See Ofcom, 2024. December 2024 Statement, Vol. 2, paragraphs 6.302 onwards; and Ofcom, 2025. April 2025 Statement, paragraphs 6.331 - 6.332.

[689] For our Protection of Children Code, services will be in scope where they are likely to be accessed by children.

[690] While we are recommending this measure due to feedback received in response to our measures proposing proactive technologies assessed by the provider, which only applies to providers of user-to-user services, we consider it is equally necessary to recommend this for search services. For example, this recommendation applies to any false positives that have been identified in quality assurance, which could take place in search moderation processes.

9.244 The full text of these measures can be found in our user-to-user and search Illegal Content and Protection of Children Codes and are referred to as ICU C4A/ICS C3A/PCU C4A/PCS C3A.

10. Hash matching

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

What is this chapter about?

Hash matching is technology used to detect known illegal or violative content. When a piece of content has been identified as being illegal, or violative, an identifier called a 'hash' (a string of identifying characters) is created to represent it and added to a database. A service then uses hash matching technology to identify content that matches the original hash. Techniques such as perceptual hashing can identify certain modified versions of content. This means the provider can either review the content or remove it automatically.

Hash matching for terror content

In this chapter we recommend the use of hash matching for terror content. We detail how this should be implemented, what constitutes an appropriate database and how human review should be utilised alongside hash matching.

Hash matching for CSAM

Our Illegal Content Codes for user-to-user services already recommend that certain service providers use hash matching to detect child sexual abuse materials (CSAM). We have decided to extend the existing hash matching measures for CSAM to include all services whose principal purpose is hosting or disseminating pornographic content, who are at high risk of image-based CSAM. We have made some changes to the provision on human review for all services to which the measure applies, to improve clarity and consistency.

Draft hash matching guidance

Alongside these measures, we have also updated our draft hash matching guidance to give service providers direction on how to implement hash matching measures.

What decisions have we made?

We have decided to proceed with both measures. In our terror hash matching measure, we have made several amendments to improve clarity. Our final Code measures are summarised in the table below:

Number in Codes	Recommended measure	Who should implement this
ICU C13	Services in scope should use hash matching technology to detect terrorism content.	User-to-user services that enable user-generated content in the form of photographs, videos and visual materials to be generated, shared or uploaded and are: a) Large services at medium or high risk of terrorism content; or b) At high risk of terrorism content and: i) have more than 700,000 monthly UK active users; or ii) are a file storage and file-sharing service.

Number in Codes	Recommended measure	Who should implement this
ICU C9	Deploy perceptual hash matching to detect known CSAM.	All high-risk services whose principal purpose is hosting or disseminating regulated pornographic content.[691]

Why have we made these decisions?

Hash matching is an effective way to identify content so that it can be moderated, or automatically taken down, depending on the type of illegal content.

Hash matching for terror content

This measure will increase the amount of terror content which service providers detect and remove. This will deliver significant benefits. Terror content is a serious and growing harm, and its availability can have severe real-world effects. Terrorism offences are often contextual; there might be legitimate reasons for sharing content related to terrorism, such as journalism. We have therefore included a number of safeguards in our measure to mitigate the risk of legitimate content being taken down inadvertently. We consider our measure strikes the correct balance in protecting users, but also ensuring that matched content is properly reviewed and, where appropriate, removed.

Hash matching for CSAM

Regardless of their size, there is a high risk of CSAM circulating on pornographic services. It is therefore essential that all pornographic services have robust safety systems, such as hash matching, in place. Our extension of the existing measure to detect CSAM will materially increase the amount of CSAM that is detected and removed from services. Given the severe and enduring damage CSAM does, this will deliver significant benefits.

Introduction

10.1 In the June 2025 Additional Safety Measures Consultation (June 2025 Consultation), we set out three proposals on the use of a proactive technology called perceptual hash matching to detect a range of illegal content types on a provider's services. Specifically, we consulted on measures recommending the use of:

- perceptual hash matching to detect terrorism content (ICU C13);
- extending the number of high-risk services in scope our existing perceptual hash matching measure for child sexual abuse material (CSAM) (ICU C9); and
- perceptual hash matching to detect intimate image abuse. We have since confirmed this measure in our May 2026 Detecting Intimate Image Abuse Statement.[692]

10.2 This chapter sets out our decision on proposals (i) and (ii). Alongside this statement, we have published the following annexes:

a) Annex 1: Further stakeholder responses

[691] Following our decisions in the December 2024 Statement on Protecting People from Illegal Harms (December 2024 Statement), this measure is already required by large services that are at medium or high risk of image-based CSAM; and by services that are at high risk of image-based CSAM and a) have more than 700,000 monthly active UK users or b) are file-storage and file-sharing services.
[692] Ofcom (2026), May 2026 Statement, Detecting Intimate Image Abuse Statement.

b) Annex 4.3: Human review for proactive technology, which includes our detailed reasoning on human review
c) Annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness
d) Annex 4.5: Hash matching for terrorism content: assessment of Tech Against Terrorism's hash list API service.

10.3 We have also published draft guidance on appropriate proportion of human review for hash matching.

Measure on perceptual hash matching for terrorism content

10.4 In our June 2025 Consultation, we proposed including a measure in our Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes) that certain providers of user-to-user services should use perceptual hash matching to detect terrorism content, where technically feasible.

10.5 We proposed the measure should apply to providers of user-to-user services that enable user-generated content in the form of photographs, videos or visual images (whether or not combined with written material) to be generated, shared or uploaded, and are:

a) large services at medium or high risk of terrorism content; or
b) at high risk of terrorism content and:
    i) have more than 700,000 monthly UK active users; or
    ii) are a file-storage and file-sharing service.

10.6 We proposed that service providers should use hash matching technology to detect known terrorism content, so that it can be removed swiftly. We proposed that service providers should compile a hash database using terrorism content identified through their own moderation processes, hashes provided by trusted external experts, or a combination of both. We said that providers would need to develop and maintain an internal database to store hashes of terrorism content removed from their service, alongside verified third-party hashes.

10.7 We said that where content is detected by hash matching, providers should treat this as a reason to suspect that the content may be illegal content and review the content to determine whether it is illegal content/illegal content proxy, in line with our content moderation measures.[693]

Our decision

10.8 We have decided to broadly confirm the measure proposed in the June 2025 Consultation, including confirming the scope of services that the measure applies to. We have made some amendments to the measure, including:

[693] Illegal content proxy: Content that a provider determines to be in breach of its terms of service, where: a) the provider had reason to suspect that the content may be illegal content; and b) the provider is satisfied that its terms of service prohibit the type of illegal content which it had reason to suspect existed. Content moderation measures: Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes), Measures ICU C1 and ICU C2.

a) Revising the criteria for defining an appropriate set of hashes. We have clarified what we mean by "persons with expertise" in the criteria for determining an appropriate set of hashes. Our reasoning is set out in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness.
b) Updating our recommendations on how services compile a hash database. The measure provides that, in certain scenarios, services should rely on a third-party database to obtain hashes of terrorism content. These scenarios, and our reasoning, are set out in the section titled 'How to source an appropriate set of hashes' in paragraph 10.21 and annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness.
c) Clarifying the position for reviewing matches against a hash database, including the factors that service providers must consider when using human review to determine whether the content is terrorism content and for the purposes of quality assurance. Our reasoning is set out in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness.
d) Replacing "visual images (whether or not combined with written material)" with "visual material". Our reasoning is set out in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness.

10.9 We are, in addition, providing draft guidance on what we consider to be an appropriate proportion of human review of detected content. This has been published alongside this statement and is titled 'Draft Guidance on appropriate proportion of human review for hash matching'.

10.10 We have also updated our position on the availability of a third-party database. In the June 2025 Consultation, we said that existing third-party databases were not appropriate for use to detect terrorism content because they are not limited to hashes of terrorism content as defined in the Act. Following further review, including assessment of additional information relating to Tech Against Terrorism (TAT), specifically its Hash List API service, we have revised our position.[694] We consider TAT to be one example of an organisation with expertise in the identification of terrorism content, and that the database provided via the Hash List API service currently meets our criteria for an appropriate set of hashes. We are satisfied that a sufficient proportion of TAT's database comprises of content amounting to UK terrorism offences, and that content which does not amount to a UK terrorism offence is either capable of being filtered out by services or presents a low risk of comprising more than a small proportion of the database.[695] Our reasoning and detailed assessment of TAT's Hash List API service is provided in annex 4.5: Hash matching for terrorism content: assessment of Tech Against Terrorism content hash list.

[694] Tech Against Terrorism (TAT) is a United Nation (UN)-launched initiative that helps global companies detect and prevent the spread of terrorist and violent extremist content online. Launched in 2016, its work involves the collection and assessment of intelligence, technical innovation, and policy advisory on counterterrorism responses. Tech Against Terrorism provides a solution, called the Terrorist Content Analytics Platform's Archive Database Hash List API. The Hash List API allows platforms to ingest hashes produced from known terrorist, violent extremist content (TVEC) media files in bulk, so they can use them in their content moderation processes. We use "database", "dataset", "hash list", and "hash list API" interchangeably in this chapter.

[695] Filtering refers to the use of predefined parameters, keywords, tags, or other criteria to select or exclude particular hashes within a hash database. In practice, this allows service providers to target specific types of content, such as hashes linked to organisations that are proscribed as terrorist organisations in the UK. In the case of TAT's Hash list API, service providers are able to select UK proscribed organisations.

How this measure works

10.11 Our measure sets out two steps that service providers should take to support the moderation of terrorism content on their service. Those two steps are:

i) detection of suspected terrorism content using perceptual hash matching; and
ii) review of detected content, to determine whether it is terrorism content, unless the provider is sure that the relevant content is illegal content (or illegal content proxy).

10.12 In accordance with existing measure ICU C2 and s.10(3)(b) of the Act, service providers should swiftly take down terrorism content which they identify.

Step 1 – Detection of suspected terrorism content using perceptual hash matching

10.13 In-scope service providers should ensure that, where technically feasible, perceptual hash matching technology is used effectively to detect suspected terrorism content. They should use the technology to analyse regulated user-generated content in the form of photographs, videos, or visual materials to assess whether it may be terrorism content.

10.14 When a service provider implements hash matching technology, the provider should consider:

a) how the hash database is compiled, including the sourcing of terrorism-related hashes and the criteria for what constitutes an appropriate set of hashes; and
b) how the hash matching technology is configured.

10.15 We explain each of these in more detail in the following paragraphs.

How to compile the hash database

10.16 Under this measure, service providers should source terrorism hashes from their own internal database or an appropriate third-party database (if they use a third- party database this should be combined with their own internal database). Both sources need to contain an appropriate set of hashes. Below, we set out the criteria for what constitutes an appropriate set of hashes and then explain the different ways in which service providers can obtain them.

Criteria for an appropriate set of hashes (database)

10.17 For a hash database to be appropriate, it should meet the following criteria:

a) Include hashes of terrorism content sourced by a person with expertise in identification of terrorism content. A "person with expertise in the identification of terrorism content" may be the service provider's content moderation staff, an expert third party database provider, or other expert in the identification of terrorism content. This person should have arrangements in place to identify suspected terrorism content and to ensure that terrorism content is correctly identified before hashes of that material are added to the database;
b) Include an appropriate range of terrorism content;
c) Contain hashes of a significant number of original items of terrorism content;
d) Not plainly discriminate on the basis of protected characteristics (within the meaning of Part 2 of the Equality Act 2010);
e) Be updated with reasonable regularity;
f) Be subject to arrangements for identifying and removing hashes that are not capable of amounting to terrorism content; and,

g) Be secured from unauthorised access or interference (whether by persons who work for that person or are providing a service to that person, or any other person).

10.18 These criteria combine our proposals on what constitutes an appropriate set of hashes with our proposals on the expectations relating to persons with relevant expertise. This differs from our approach in the consultation, where we presented separate recommendations for "an appropriate set of hashes" and for "persons with expertise".[696] We have combined them here to improve clarity and simplify the measure.[697] Our reasoning for updating the criteria is set out in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness.

10.19 Where the provider becomes aware of terrorism content, which is not included in its hash database, the provider should ensure that a hash of that content is added to its database. This may occur, for example, where the provider identifies terrorism content through its moderation processes.

10.20 Providers should ensure an appropriate policy is put in place, and measures are taken in accordance with that policy, to secure any hashes from unauthorised access, interference or exploitation.

How to source an appropriate set of hashes

10.21 There are two main ways for service providers to obtain an appropriate set of hashes:

a) **In-house content**: Service providers can source hashes from terrorism content previously identified on their own services, where the provider still has access to the content.[698] However, this approach will only be sufficient where the resulting database meets our criteria of an appropriate set of hashes, as set out in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness. This includes but is not limited to, that it contains hashes of a significant number of original items of terrorism content.[699]

b) **Combination (in-house and third party)**: Where a provider's database does not meet these criteria, it will also need to supplement its dataset with hashes from a third-party provider whose dataset meets our criteria.

10.22 Regardless of the approach taken, whether developing the database fully in-house or a combination of in-house and third-party sources, the resulting database must meet the criteria set out in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness. We expect service providers to implement the measure as soon as reasonably practicable.

10.23 We consider that TAT's Hash List API service is an example of a third-party database that currently meets our criteria for an appropriate set of hashes. Our detailed assessment of TAT's Hash List API service is set out in annex 4.5: Hash matching for terrorism content:

---

[696] Ofcom (2025), June 2025 Additional Safety Measures Consultation, p. 153-154.

[697] The approach we take here is different from the approach we have taken to CSAM hash matching and to intimate image abuse hash matching because we consider that some service providers may have internal databases of sufficient size and range to be effective, and because we do not consider that use of an unverified database would be effective in the case of the terrorism offences.

[698] Illegal Content Codes, Measures ICU C1 and C2.

[699] We would expect this to comprise at least the number of original items of content that the provider can source using a third party such as TAT.

assessment of Tech Against Terrorism content hash list. Other organisations or providers may also meet these criteria, and it is for service providers to assess this.

Configuration of the hash matching technology

10.24 Service providers within scope of the measure should deploy hash matching technology to detect terrorism content before, or as soon as practicable after, such content can be encountered by UK users. The hash matching technology should use a suitable perceptual hash function to compare relevant content to an appropriate set of hashes.

10.25 Content already on the service when the hash matching technology is implemented should be moderated within a reasonable time frame.[700]

Striking an appropriate balance between precision and recall

10.26 Service providers should ensure the hash matching technology is configured to strike an appropriate balance between precision and recall.[701] In doing so, service providers should consider the following factors:

a) risk of harm to users from relating to terrorism, reflecting the service's latest illegal content risk assessment, and any information reasonably available to the provider about the prevalence of terrorism content;
b) the proportion of content detected as a match by the hash matching technology that is a false positive (for example, where the image is not capable of being terrorism content); and
c) effectiveness of the systems and processes used to identify false positives (e.g. user complaint mechanisms).[702]

Reviewing the balance between precision and recall

10.27 Service providers should review the balance between precision and recall at least every six months, to ensure that it remains appropriate.

10.28 Service providers should keep a written record of their approach to configuring the technology, including what information has been considered, and information about reviews and steps taken in response.

Analysis of relevant content

10.29 The service provider should ensure that all relevant content present on the service at the time the perceptual hash matching technology is implemented is analysed within a reasonable timeframe.[703]

10.30 The service provider should also ensure that any relevant content generated, uploaded or shared on the service thereafter is analysed before it can be encountered by UK users or, where that is not practicable, as soon as reasonably practicable thereafter.

---

[700] Illegal Content Codes, Measures ICU C1 and C2.
[701] Precision is the proportion of identified cases that are true positives. Recall is the proportion of true positive cases that are correctly identified.
[702] We consider our existing user appeals measures in the Illegal Content Codes Measures ICU D1, D2, D8, D9 and D10 will facilitate appropriate mechanisms for users to submit complaints to service providers if their content has been wrongly removed.
[703] Relevant content: "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 by means of the service, or (b) any material which, if it were present on the service, would be content within sub-paragraph (a).

Step 2 – Review of detected content, to determine whether it is terrorism content (or illegal content proxy).

10.31 Where relevant content matches with a hash, the provider has two options. When the provider has a high level of assurance in the outcomes achieved by hash matching technology, content should be treated as illegal and removed swiftly in line with ICU C2.3 (Having a content moderation function that allows for the swift take down of illegal content).[704] Where this is not the case, content matching with a hash means the provider has reason to suspect that the relevant content may be illegal. The content should then be reviewed in accordance with ICU C1 (Having a content moderation function to review and assess suspected illegal content).[705]

10.32 Where a review is needed, it may be carried out through automated processes or by human reviewers. However, we do not anticipate that there will be many scenarios in which the provider can have a sufficient level of assurance in the detection outcomes achieved by technology to be able to do without human review, because many of the terrorism offences in UK law are context specific. This means that the same content may be illegal content in some contexts, and not in others.[706] It is therefore not necessarily possible to infer that matched content is illegal content without further context. It is likely that services will be relying on human review.

10.33 For the purposes of this measure, human review can be used either:

a) to make decisions about individual items of content (content moderation); and
b) to test, monitor and improve the accuracy of moderation tooling and processes (for quality assurance).

10.34 We discuss each of these forms of human review in more detail in annex 4.3: Human review for proactive technology. We have also issued draft guidance to assist providers in applying this part of the measure (Draft Guidance on appropriate proportion of human review for hash matching). The draft guidance provides more detail on human review and what factors service providers should consider.

Human review for content moderation purposes

10.35 Human review for content moderation purposes under this measure refers to the use of human moderators to review an appropriate proportion of all detected content to determine whether a piece of content is illegal and should be swiftly removed.

10.36 The provider should ensure human moderators review and assess an appropriate proportion of detected content, having regard:

a) to the level of assurance the provider has in the detection outcomes achieved by the hash matching technology and any associated systems and processes, as indicated by the ongoing monitoring, evaluation and quality assurance (including human quality assurance) of the performance of the technology; and
b) to the extent that detected content is subject to review for the purpose of Recommendation ICU C1:

[704] Illegal Content Codes, Measures ICU C2.3.
[705] Illegal Content Codes, Measures ICU C1.
[706] For example, a video of a terror attack may be circulated as part of news reporting or for academic research, or it may be distributed by a terrorist group to encourage further attacks.

i) the potential severity of the harm to those depicted in or encountering terrorism content; and
ii) the overall impact of an incorrect decision that the relevant content is illegal content on a UK user who generated, uploaded or shared the content.

10.37 Where the provider's level of assurance in the detection outcome is high, for example if multiple detection technologies are layered, it may be that reviews can be handled through automation. If detection relies primarily on hash matching and confidence in accuracy is limited due to a high number of false positives, a higher proportion of detected content should be escalated to human review. Human review for quality assurance purposes will still be needed.

10.38 In limited cases, for example where the content is clearly information likely to be of use to a terrorist or amounts to a terrorist training offence, there may be suspected terrorism content which does not require additional context to make an illegal content judgement. In such cases, providers may choose to rely on automated review rather than human review, depending on the level of assurance they have in the detection outcomes achieved by the technology. Section 2 of our Illegal Content Judgments Guidance (ICJG) explains how to make illegal content judgements.[707]

Human review for quality assurance purposes

10.39 Human review for quality assurance purposes under this measure refers to the use of human moderators to review an appropriate proportion of all detected content to identify any systemic issues and to improve the hash matching technology.

10.40 Service providers should use human review for the purposes of quality assurance and monitoring of the configuration of technology. This review will apply to any technology used for the purposes of this measure, including hash matching technology and other automated technologies (these may be used for the review of content, as set out below).

10.41 This element of the measure reflects amendments to the factors that service providers should take into account when determining what proportion of detected content should be subject to human review. We explain our reasoning for amending the factors in annex 4.3: Human review for proactive technology. We are also providing additional draft guidance to assist providers.

10.42 As mentioned in chapter 9: Deployment of proactive technology to detect image-based CSAM – Steps to take on reversal of content decisions, we have added a measure to our content moderation measures a safeguard for freedom of expression in all our Codes of Practice to ensure that decisions are reversed so far as possible and appropriate where providers find them to be false positives.[708] We expect this to be relevant principally when a provider carries out quality assurance, but it is possible that a provider might learn of an error in other ways (for example, if notified of a court decision).

Appropriate actions based on the content review outcome

10.43 When a service provider identifies new terrorism content on its service, it should add hashes of this content to its internal database and take down the content. Illegal content, or illegal content proxy, should be swiftly taken down in accordance with measure ICU C2

[707] Ofcom (2026), Illegal Content Judgements Guidance.
[708] Chapter 9: Deployment of proactive technology to detect image-based CSAM – Steps to take on reversal of content decisions. Measure ICU C4A.

(Having a content moderation function that allows for the swift take down of illegal content).[709]

If a service provider becomes aware of a hash of content that is not capable of amounting to terrorism content

10.44 The measure sets out that a service provider should take reasonable steps to secure the removal of a hash from the hash database where it has determined that the hash is not capable of amounting to terrorism content. Hashes that may correctly match content capable of amounting to terrorism content in some circumstances should not be removed and should continue to be detected and reviewed. A provider may determine that a hash is not capable of amounting to terrorism content in the following circumstances:

a) the provider has viewed the content used to generate the hash and determined it is not capable of amounting to terrorism content;
b) the provider has viewed content matched to the hash using perceptual hash matching technology and determined that the content is not capable of amounting to terrorism content. This applies only where the provider reasonably believes that the hash of the detected content exactly matches the hash contained in the database.

10.45 When a service provider uses an internal database, taking reasonable steps to secure the removal of the hash would involve removing the hash from that database. If a provider is using a third-party database, it will satisfy this component of the measure by notifying the hash database provider of the incorrect hash.

10.46 For clarity, our measure does not impose recommendations on third-party database providers to remove hashes from their database that are determined not to be terrorism content; we do not have any power to regulate providers of third-party databases in this regard. However, we encourage third-party database providers to remove the incorrect hashes that service providers report, to ensure greater accuracy of their database.

Interaction with recommender systems recommendation

10.47 Content for which a hash match has been detected should, at least, be considered content suspected to be illegal terrorism content. Where this content is sent for human review, we consider it should be excluded from content recommender feeds of children under our recommender systems measures (ICU E2). We consider information available from the service's hash matching systems and processes to be sufficient "relevant available information".

10.48 Services should refer to chapter 11: Recommender systems for further information.

Our reasoning

10.49 In this section, we detail our reasoning for including this measure in our Illegal Content Codes. This includes an analysis of stakeholder feedback on the following topics:

a) effectiveness and benefits of the measure;
b) costs and impacts of the measure; and
c) impact on users' rights from the measure.

[709] Illegal Content Codes, Measures ICU C2.

Effectiveness and Benefits

10.50 The primary benefit of the measure is to reduce the generation, upload or sharing of terrorism content, and, by extension, the harm caused by the content.[710] This is done via the use of hash matching. As explained in chapter 8: Introduction to proactive technology, the Act requires us to consider the accuracy, effectiveness, and lack of bias of proactive technology that we include in our Codes of Practice.

10.51 In this section, we demonstrate how we have considered this by first discussing the harm terrorism content causes, and the effectiveness of hash matching technology to identify this content. This includes its limitations (such as the availability and accuracy of databases) and the safeguards available to minimise their impact. It also considers how this technology minimises the risk of bias.

Terrorism content causes serious harm

10.52 Online services can enable terrorist organisations and their supporters to reach large audiences and distribute content quickly and widely.[711] It is a growing harm.[712]

10.53 Access to the internet has enabled the spread of terrorism content. As the United Nations has noted, "almost all terrorist action is at least partly reliant on internet communication. Through the web, violent messages are spread, contact between perpetrators is facilitated, and attacks are incited. Websites, social media and instant messaging are widely used by terrorists to radicalise and recruit individuals and to plan, fund or commit terrorist attacks".[713]

10.54 Terrorist actors use a range of online services, including forums, chatrooms, instant messaging services and mainstream social media services, to communicate, recruit, disseminate propaganda and support operational activity.[714] Research has found that online communities can play an important role in processes of radicalisation towards extremist offending, and that mainstream social media and file-sharing services have experienced extremist activity to varying degree.[715] The internet is evidently recognised as a key environment in which terrorism ideologies are disseminated, supporters are mobilised and attacks may be planned or encouraged. According to the Government's interim Code, "almost all UK terrorist attack planners from 2012-2017 downloaded, shared or consumed content and activity associated with terrorism and extremism online".[716]

10.55 Our research shows that while 11% of adult users had encountered extremist content which they perceived to be illegal, 58% of the sample reported that this encounter had a negative impact on them.[717]

---

[710] Section 3 of the Act.
[711] UK Parliament, 2020. Postnote - Online Extremism. Number 622. Home Office, 2020. Interim Code of Practice on Terrorism Content and Activity Online December.
[712] Home Office, 2020. Interim Code of Practice on Terrorism Content and Activity Online December.
[713] United Nations Office on Drugs and Crime (n.d.) The use of the Internet.
[714] Kenyon, J., Binder, J. and Baker-Beall, C. 2022. The Internet and radicalisation pathways: technological advances, relevance of mental health and role of attackers. Ministry of Justice Analytical Series. London: Ministry of Justice, HM Prison & Probation Service, p.6.; Winter, C., Neumann, P., Meleagrou-Hitchens, A., Ranstorp, M., Vidino, L., & Fürst, Johanna. 2020. Online extremism: Research trends in Internet activism, radicalization, and counter-strategies. International Journal of Conflict and Violence, 14(2), 1–20.
[715] Kenyon, Binder and Baker-Beall, The Internet and radicalisation pathways.
[716] Home Office, 2020. Interim Code of Practice on Terrorism Content and Activity Online December.
[717] Ofcom (2025), Online Nation Report 2025, p.70.

10.56 Beyond these impacts, widespread availability of terrorism content can contribute to real-world harm. In the most serious cases, exposure to this content can contribute to attacks that result in serious injury or loss of life.[718]

Effective handling of user reporting and complaints is not sufficient to address terrorism content and its harm

10.57 We consider that the use of some form of proactive technology is necessary to protect users against terrorism content. Reliance solely on user complaints and reports is unlikely to be sufficient, as such approaches are inherently reactive and may not enable service providers to identify and address terrorist content quickly enough to mitigate harm. Given the high volume of content that is shared, uploaded and generated on user-to-user services, proactive technology such as hash matching is needed to detect the presence of terrorism content. For large services in particular, we understand that millions of content items are uploaded every day.[719] We discuss the use of proactive technology in more detail in chapter 8: Introduction to proactive technology.

10.58 Terrorism offences can be complex and context-dependent, and they may be difficult for users to identify without specialist knowledge.[720] Even where users recognise terrorism content, they may not be willing or motivated to report it. Content may circulate within communities where users share ideological alignment or a common interest in not reporting. In such cases, there may be a strong disincentive to report content that supports or promotes the group's aims, even where it is clearly illegal.

10.59 Children and young people are unlikely to report content, particularly where it is disturbing, confusing, or normalised within their online environment.[721] This limits the effectiveness of reporting-based approaches, especially given that terrorism propaganda and recruitment narratives are sometimes designed to appeal to younger audiences.[722] Therefore, we have concluded that existing user reporting and complaints mechanisms are not sufficient on their own to protect users from terrorism content.

---

[718] Fetiu, A. and Alava, S. (2018) "Exposure to Extremist Online Content Could Lead to Violent Radicalization: A Systematic Review of Empirical Evidence," International Journal of Developmental Science, doi: 10.3233/DEV-170233.

[719] Publicly available figures from providers of large services. For example, Reddit reported that, globally, approximately 260 million posts and 2 billion comments were created in H2 2025, equivalent to around 1.4 million posts and 11 million comments per day. Reddit, 2026. Transparency Report; July to December 2025. Snap reported that in 2025, global users on average posted 5.5 billion Snaps per day. Snap Inc, 2026. Snapchatters Created Nearly 2 Trillion Snaps in 2025.

[720] This is evident when considering the definition of terrorism as described in chapter 2 of the Illegal Content Judgements Guidance, particularly with reference to paragraphs 2.8 to 2.15. As of January 2026, 84 terrorist organisations are proscribed under the Terrorism Act 2000, 14 organisations in Northern Ireland that were proscribed under previous legislation. Home Office, 2026. Proscribed terrorist groups or organisations. Under UK law, there are over 80 proscribed terrorist organisations, including both international and UK based groups, many of which may be unfamiliar to the general public. Terrorism content under the Act is defined by reference to existing UK terrorism offences and covers a wide range of offences, such as expressing an opinion or belief supportive of a proscribed organisation, inviting support for a proscribed organisation, training, fundraising, and the dissemination of terrorist publications. Some of this content may appear ambiguous, or be difficult to distinguish from lawful material, for an untrained user.
Ofcom (2025), Online Nation Report 2025, p.10.

[721] Ofcom (2025), Online Nation Report 2025, p.10.

[722] Rose, H. and Vale, G. 2023. Childhood Innocence?: Mapping Trends in Teenage Terrorism Offenders. London: International Centre for the Study of Radicalisation (ICSR), King's College London, p.47.

Hash matching is effective and accurate at detecting terrorism content

10.60 We have carried out a more detailed assessment of the effectiveness and accuracy of perceptual hash matching for terrorism content in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness. The principal stakeholder feedback and assessment findings that underpin our conclusion that perceptual hash matching is an effective and accurate technology for detecting terrorism content are summarised below.

Hash matching is effective because it supports greater scale and broader coverage in identifying terrorism content

10.61 In response to our June 2025 Consultation, several respondents noted that hash matching can be effective at identifying known content.[723] [✂].[724] [✂].[725]

10.62 We consider that hash matching technology supports greater scale and broader coverage in identifying terrorism content because it enables service providers to compare content against databases containing large volumes of known terrorism content. This can help providers identify content that may otherwise be difficult to detect, improving the consistency and breadth of content identification across a service.

Context may impact the accuracy of hash matching

10.63 We recognise that because many terrorism offences are contextual,726 the case for terror hash matching is more nuanced than it is for less-contextual offences like CSAM.

10.64 Several stakeholders suggested that terrorism content is context dependent.727 Some respondents suggested that hash matching can be effective at detecting some categories of content such as CSAM, but less effective at detecting other types of harms such as terrorism content where context is needed.728

10.65 In our view, this does not mean that hash matching is inappropriate for detecting terrorism content. First, some terrorism content can be identified as such without extensive contextual analysis. Second, hash matching is an effective means of identifying content that there is reason to suspect may be terrorist in nature. For example, it is possible to communicate a past call to murder for news-reporting or academic purposes, but its circulation within communities sharing violent ideologies may amount to a renewal of the call to murder. Where such content is detected, it is appropriate for service providers to assess it through moderation and determine whether or not it amounts to terrorist content and should therefore be removed. Given the seriousness of terrorism-related harms, using hash matching in this way can play an important role in enabling services to identify potentially terrorist content.

10.66 As context may be relevant to the final moderation outcome, service providers will need to ensure that detected content is subject to further review, whether through automated

---

[723] Microsoft response to the June 2025 Consultation, p.8; [✂].
[724] [✂].
[725] [✂].
[726] By "contextual", we mean that to determine whether the content amounts to an offence, the provider is likely to need to consider information other than the specific item of content concerned. The extra information needed varies from offence to offence but could include the state of mind of the person posting the content, its meaning in light of the content around it and who is likely to see the content. Guidance on how to make these judgments is set out in Ofcom's Illegal content Judgments Guidance.
[727] British and Irish Law Education and Technology Association (BILETA) response to the June 2025 Consultation, pp.18-19; Google LLC response to the June 2025 Consultation, p.21 -22]; [✂]; [✂]; [✂].
[728] Google LLC response to the June 2025 Consultation, p.7; [✂].

systems, human review, or both. In practice, we expect that the level of assurance in the detection outcomes produced by the hash matching technology will be low for the majority of the terrorism offences. This is because most terrorism offences are contextual, requiring providers to consider the context of the content, including any relevant intent of the posting user, when determining whether the content is illegal.

10.67 At the time of issue of this statement, Ofcom was not aware of any available technology that could reliably make such contextual assessments without human input. Therefore, human review is likely for most hash matches. We discuss this relationship between detection and review further in annex 4.3: Human review for proactive technology, and in the draft Guidance on appropriate proportion of human review for hash matching, published alongside this statement.

The effectiveness of hash matching depends heavily on the quality of the underlying hash database

10.68 Our assessment found that the effectiveness of hash matching depends heavily on the quality of the underlying hash database. For this reason, we have set out criteria in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness for what constitutes an appropriate set of hashes, including recommendations relating to the breadth of terrorism content covered, the scale of the dataset, regular updates, security, and processes for identifying and removing inaccurate hashes. Recognising the time it is likely to take for service providers to build up an effective database solely through content moderation, we have also concluded that providers without an appropriate internal dataset should supplement their dataset with a suitable third-party database.

10.69 We consider that at least one third-party database meets the criteria relating to database quality. We have an assessment of the hash database provided by TAT as an example of a database that meets those criteria in annex 4.5: Hash matching for terrorism content: assessment of Tech Against Terrorism's content hash list API service.

Perceptual hash matching is an effective and accurate means of detecting known terrorism content

10.70 We recognise that perceptual hash matching is not effective at identifying entirely new or previously unseen content, as there is no existing material to compare it against. However, the fact that perceptual hash matching cannot detect entirely new terrorism content does not mean it is ineffective. It remains a highly effective means of identifying known terrorism content and can make a significant contribution to reducing users' exposure to such content.

10.71 We also consider that perceptual hash matching is significantly more effective than cryptographic hash matching for identifying and preventing the recirculation of known terrorism content, as it can detect not only exact copies but also content that has been subject to minor modifications. This makes it more resilient to common evasion techniques and enables the detection of a wider range of terrorism content.

10.72 We have received feedback on the risk of bias when using proactive technology, specifically hash matching technology.[729] We have not identified significant risks of algorithmic bias arising from hash matching technology itself, given its deterministic nature. The elements of the measure relating to quality assurance and human review are intended to help providers

[729] BILETA response to the June 2025 Consultation p.15; Open Rights Group response to the June 2025 Consultation, p.27; Digital Resilience in Education – Welsh Government response to the June 2025 Consultation, p 10.

identify and address any such issues in practice. The decisions we have taken on the detail of the design of the measure are intended to maximise the efficacy of the measure.

Accurate and effective deployment of hash matching technology requires service providers to appropriately configure perceptual hash matching systems and regularly review their performance

10.73 Finally, our analysis indicates that accurate and effective deployment, that is lacking in bias, requires providers to appropriately configure perceptual hash matching systems and regularly review their performance. The technology is effective when supported by a high-quality, well-maintained dataset and when providers calibrate their systems to achieve an appropriate balance between detecting terrorism content and minimising false positives.

10.74 We have carried out a more detailed assessment of the effectiveness and accuracy of perceptual hash matching for terrorism content in annex 4.4: Hash matching for terrorism content: detailed reasoning on effectiveness.

An appropriate proportion of human review is necessary to ensure accuracy and effectiveness, and to minimise the risk of bias

10.75 We have updated our position on the factors that should be considered when determining what constitutes an appropriate proportion of human review. Our analysis suggests that for hash matching to be effective, accurate and to lack bias, an appropriate proportion of human review is necessary. We set this out in annex 4.3: Human review for proactive technology and provide further guidance in our draft Guidance on appropriate proportion of human review for hash matching (published alongside this statement).

Conclusion on effectiveness and benefits

10.76 The analysis above shows that terrorist content is widespread online and has a significant adverse effect. It also shows that perceptual hash matching technology is an effective tool for combatting the spread of such content and that the benefits of deploying this technology cannot be replicated by user reporting alone. Therefore, we have concluded that the use of perceptual hash matching technology will reduce the spread of illegal material online, thereby materially lowering the risk of radicalisation; limiting terrorist groups' ability to disseminate propaganda; and mitigating the ongoing harm and fear intended through the circulation of such content.

10.77 We consider that perceptual hash matching remains an appropriate means of identifying suspected terrorism content, that at least one suitable hash database is available to support its deployment, and that the safeguards built into the measure, including human review and quality assurance processes, help mitigate the risk of legitimate content being incorrectly actioned.

10.78 Taking these factors together, we consider that the measure is effective means of reducing the spread of terrorism content online. We consider that the technology, when deployed as recommended, is accurate, effective, and minimises the risk of bias.

Impacts and costs

10.79 In our June 2025 Consultation, we outlined our assessment of costs for the implementation of hash matching for terrorism content. We noted that the costs incurred by service providers would vary according to a service's number of users, technical complexity, and risk of terrorism content.

10.80 Some stakeholders recognised similar cost considerations to those identified in our assessment.730 Other stakeholders disagree with our assessment of the costs of this measure.731 The Middle Tech Coalition raised concerns about the costs associated with implementing and maintaining general perceptual hash matching technologies, noting that these can vary significantly between services.732 However, no stakeholders presented evidence which warrants fundamentally revising our cost analysis.

10.81 As set out in the consultation, the costs of the measure are likely to consist of the following components: costs of building and maintaining a hash matching system; costs of buying software, hardware and data; and costs of reviewing and moderating content. We discuss each of these cost categories below.

Cost of building a hash matching system

10.82 A service provider will need to build a system that can run hash matching tools and store hashes of illegal content in an internal database which will involve labour costs. In our June 2025 Consultation, we said if a service already uses hash matching for CSAM, then costs may be lower due to the sharing of supporting technology for both harms although this might not always be the case.

10.83 Some stakeholders said that implementing hash matching for terrorism would be costly even if a service already used hash matching for other harms such as CSAM.733

10.84 We maintain that some service providers who use hash matching for other harms may face lower costs when implementing hash matching for terrorism content. This will depend on the hashing tools and functions already in use as well as the nature of the service. In cases where the technologies used by service providers cannot be simply or uniformly applied to terrorism content, this will result in the provider incurring the full costs of implementing the measure.

10.85 We have updated our quantified estimates to bring them in line with the latest ONS salary data.[734] We estimate a service provider which does not already use hashing technologies will incur a one-off cost of around £22,000 to £387,000 to build a hash matching system for terrorism content.[735] We expect that a large proportion of in-scope service providers would incur implementation costs in line with the lower end of this estimated cost range, with costs increasing with user numbers and service complexity.

Cost of maintaining a hash matching system

10.86 A service provider will need to maintain its hash matching system to ensure it is up to date; which may include adding or removing hashes from the internal database. This means there will be on-going maintenance and operational costs of software engineering time.

---

730 BILETA response to the June 2025 Consultation, p.17; [✂].
731 [✂]; [✂]; Mega Privacy LLC response to the June 2025 Consultation, p.18; [✂]; Pinterest response to the June 2025 Consultation, p.5; [✂]; [✂].
732 Middle Tech Coalition response to the June 2025 Consultation, p.4.
733 CCIA response to the June 2025 Consultation, p.10; [✂].
734 See annex 5: Further detail on economic assumptions and analysis.
735 This is based on the cost of a software engineer working two to eighteen months with an equal amount of professional occupation staff time.

10.87 Some stakeholders raised concerns about the cost of maintaining an internal hash database of terrorism content.736

10.88 The measure now provides for service providers to use a third-party hash database where they do not already use hash matching or in cases where their internal dataset does not consist of an appropriate set of hashes. This will help reduce the labour cost of maintaining an effective hash database as providers can immediately access a substantial volume of hashes that have already been identified on other services. Using an external database may lead to additional costs of subscribing to an appropriate third-party hash list. We discuss this under the section titled 'Costs of software, hardware and data'.

10.89 We estimate that a service provider will incur an annual cost of around £5,000 to £97,000 to maintain its hash matching system.[737] As with the build costs we expect service providers to incur annual costs relating to hash matching system maintenance in proportion to service size and complexity. It is likely that small services will incur costs towards the lower end of this range.

Cost of software, hardware and data

10.90 In our June 2025 Consultation, we outlined that there will also be non-labour related costs incurred from implementing the measure. These costs include the hardware and software needed to generate hashes, and the systems and processing required to match content against existing hashes. In the cases where service providers opt to use a third-party database of hashes,738 it includes the cost of acquiring or referencing this third-party database.

10.91 In our June 2025 Consultation, we provided an indicative estimate for the annual cost of software, hardware and data of £100,000 to £150,000 for a large service that would need to hash up to eighty-seven million pieces of content per day.739 If a provider uses cloud services for hash matching, there may be overlap between the costs of software or hardware, and the cost category for building a hash matching system. As most services would need to hash significantly less than eighty-seven million pieces of content per day, this estimate is representative of what a large service provider would incur.

10.92 Based on evidence from the Tech Coalition, we recognise that some forms of hash matching could be deployed by smaller services using conventional and non-specialist operating systems or software, while achieving a satisfactory benchmark of performance.[740] In addition, supporting evidence from AiLECS suggests that hash matching could be performed locally or at edge level as opposed to remotely.[741] Overall, the precise costs will depend on the service's implementation choices, risk profile and the speed at which content needs to

---

[736] BILETA response to the June 2025 Consultation, p.17; Online Safety Network response to the June 2025 Consultation, p.18; Tech Against Terrorism response to the June 2025 Consultation, p.9.
[737] This is based on our standard assumption that ongoing maintenance costs are 25% of the one-off build cost incurred in year 1. See annex 5: Further detail on our economic assumptions and analysis.
[738] Hasher-Matcher-Actioner is one example of a free to use and open-source third-party software. Services could use such a tool to implement hash matching for terrorism content.
[739] We modelled a theoretical peak performance of 1500-2000 queries per second with 50% redundancy and latency of less than 10ms per query. This would mean 87 million requests per day which we estimate would require £100,000 to £150,000 of either infrastructure or rented cloud computing capability.
[740] Morales, F., Sharma, S., Doan, C. and Levine, B. .2024. PHVSpec: A benchmark-based analysis of perceptual hash systems for videos. Technology Coalition,Table 3, p.14.
[741] Dalins, J., Wilson, C. and Boudry, D. 2019. PDQ & TMK + PDQF: A Test Drive of Facebook's Perceptual Hashing Algorithms. arXiv. P.13.

be detected and reviewed. On this basis, we expect software and hardware costs for small services to be significantly lower than £150,000 depending on content volumes and the risk of harm on the service.

10.93 The measure now provides for some services to use a third-party hash list when developing their database which means these services will need to incur the associated membership or subscription fees to access these third-party hash lists. Some stakeholders said that accessing third-party hash lists is costly.[742] [✂].[743]

10.94 We consider TAT's Hash List API service to be an example of a third-party dataset that can meet our criteria for an appropriate set of hashes and can be used by services to develop their database. Access to TAT's hash list involves an annual subscription. Access is tiered by global monthly active users. Services under 10 million may be eligible for free access, services from 10 million to under 45 million are eligible for discounted access, [✂], and services of 45 million or more gain access through a commercial agreement [✂].[744]

10.95 Service providers which continue using a third-party dataset would incur repeated costs of subscription. Other providers may find an initial seeding from a third-party dataset to supplement their internal dataset, before relying wholly on their own reporting and moderation processes, to be sufficient to maintain an effective database. These providers would not have annual subscription costs though they may incur higher costs of reviewing user reports of terrorism content and adding hashes to their database, which we account for in the cost of maintaining a hash matching system.

Cost of review and moderation

10.96 Matched hashes need to go through content moderation, including an appropriate proportion of human review. This has cost impacts for services.

10.97 In our June 2025 Consultation, we provided estimates for the annual cost of human moderation on a hypothetical large service which needs to review 2,600 pieces of content a year following a match.745 We said that the cost incurred by providers would increase with the number of matches and therefore the volume of content on a service.

10.98 We received some feedback relating to the cost of moderation, including the training of moderators. We address this feedback in the relevant chapter in annex 1: Further stakeholder responses.

10.99 We have assumed in our cost estimate that a human moderator reviews every piece of content flagged by the hash matching system. This is so that we do not underestimate the degree of human review for a given volume of content. However, in practice providers may only need human review for a proportion of content. The proportion of detected content subject to human review should be determined by considering the factors that we are recommending, as discussed in annex 4.3: Human review for proactive technology.

10.100 We have updated our estimate according to the latest salary data and estimate that a provider of a hypothetical large service would incur costs of moderation of around £16,000

---

[742] [✂]; Pinterest response to the June 2025 Consultation, p.5.
[743] [✂].
[744] Tech Against Terrorism, 2026. Hash List Access Tiers | Terrorist Content Analytics Platform.
[745] This is based on a hypothetical service with 70 million global users of which 7 million are in the UK. Our estimate assumes there are 0.00004 items of terrorism content uploaded per user and 0.0013 moderators are needed per piece of content actioned.

to £32,000 per year. We expect that smaller services will likely incur costs that are relatively low and possibly lower than the provided range.

Conclusion on costs and impacts

10.101 We summarise the cost to service providers in Table 10.1. We estimate that providers of smaller services will incur costs towards the lower end of the ranges. If a service also uses hash matching to detect CSAM or intimate image abuse, it may incur reduced costs as there could be overlaps in technology.

Table 10.1: Estimated costs to service providers of implementing hash matching for terrorism content

Component	Estimated cost
Build cost (one-off)	£22,000 to £387,000
Maintenance costs (annual)	£5,000 to £97,000
Software, hardware and data (annual)	Up to £150,000 for a large service, plus the cost of a third-party hash list if used. Significantly lower for small services.
Human moderation (annual)	£16,000 to £32,000 for a large service.

Rights impact

Freedom of expression

10.102 Stakeholders raised concerns that the proposed use of hash matching and other automated moderation technologies could negatively impact freedom of expression including through the over removal and suppression of lawful and legitimate content.[746] Concerns from stakeholders included the issue of false positives and its impact on freedom of expression,[747] lack of contextual assessment,[748] potential bias against certain groups,[749] and the need for human review.[750] Some stakeholders also questioned the clarity and foreseeability of the proposals.[751]

Is there an interference?

10.103 We accept that the measure we are recommending could interfere with the right to freedom of expression. It is linked to our wider suite of content moderation Codes of

[746] [✂]; [✂]; The Open Rights Group Campaign response to the June 2025 Consultation, pp.4-5; Open Rights Group response to the June 2025 Consultation, pp. 24-31; [✂]; [✂].
[747] BILETA response to the June 2025 Consultation, pp 16; [✂]; [✂]; [✂]; Open Rights Group response to the June 2025 Consultation, pp. 24-31.
[748] [✂]; [✂]; Open Rights Group response to the June 2025 Consultation, p27. [✂].
[749] BILETA response to the June 2025 Consultation p.15 p.6; Open Rights Group response to the June 2025 Consultation, p.27; Digital Resilience in Education - Welsh Government response to the June 2025 Consultation, p 10.
[750] BILETA response to the June 2025 Consultation, pp 16-17; Computer and Communications Industry Association response to the June 2025 Consultation: p10; Evangelic Alliance response to the June 2025 Consultation, p.6; [✂]; Open Rights Group response to the June 2025 Consultation, p.25-29.
[751] Open Rights Group response to the June 2025 Consultation, pp.24-31, the Open Rights Group Campaign response to the June 2025 Consultation, pp.4-5; Cyberleagle, 2025. Ofcom's proactive technology measures: principles-based or vague?.

Practice measures which ultimately lead to the take down or removal of content. We have considered the impact of our measure in this context.

10.104 The measure operates alongside other elements of the regulatory framework. In particular, relevant available information generated through hash matching may be used to inform recommender systems, including by limiting the visibility or distribution of content pending assessment. It also interacts with existing content moderation measures, which provide for the review and, where appropriate, the take down of illegal content. We consider the potential impact on users' rights arising from these interactions, including their combined effect, in this chapter and further in chapter 11: Recommender systems.

10.105 Terrorism content for the purposes of this measure, is content which amounts to an offence as specified in Schedule 5 to the Act, including but not limited to offences relating to proscribed organisations, encouraging terrorism offences, training and financing terrorism. Parliament has defined terrorism content as priority illegal content for the purposes of the Act. It has required that proportionate steps be taken to prevent users from encountering this type of content and to ensure it is taken down or removed when service providers become aware of it. As such, to the extent that the measure impacts on terrorism content, we consider this flows from the requirements of the Act.

10.106 However, the way that providers implement the measure may also affect legal content. An interference with users' freedom of expression is likely to arise when lawful content is detected and is taken down or removed as a result.[752]

10.107 Incorrect decisions arising from use of a hash matching database may arise in several ways:

a) Hashes of non-terrorism images and videos may be included in a database, either in error or maliciously;
b) Content may be correctly included in the database, but may only amount to terrorism content in certain contexts, and absent safeguards will therefore potentially generate an incorrect decision wherever the required context is not present; or
c) Hashes of terrorism content may be included on a database, but may match with hashes of lawful content due to the configuration of the hash matching parameters.

10.108 As we have discussed in detail in the section titled 'Effectiveness and benefits', the nature of terror content might include what would otherwise be the most highly protected forms of speech – political speech and sometimes also religious speech. Users likely to post content that is similar to known terrorist content are particularly likely to be affected, both by the risk of error and by the risk that context may not be sufficiently taken into account. We especially noted that activists documenting legitimate issues may be inadvertently caught by perceptual hash matches that fail to account for context.[753]

---

[752] In principle, there could also be a more general 'chilling effect' if some users were to avoid using services which had implemented the measure due to concerns about adverse impacts on users' rights (as noted by Open Rights Group in its response to our June 2025 Consultation, p.24). However, many UK users already use services which have implemented perceptual hash matching. Any such effect therefore seems unlikely to be significant. Similarly, there could be a chilling effect if providers ceased to provide services in the UK or geo-blocked content so that it cannot be encountered in the UK as a result of the measure. As we know that many providers are already implementing hash matching, and UK users in any event have many thousands of public and private services to choose from, we do not consider such impacts to be likely or significant.

[753] Ofcom, (2025), paragraph 12.69, citing Electronic Frontier Foundation, 2019: Caught in the Net: The impact of "extremist" speech regulations on human rights content.

Prescribed by law

10.109 To be prescribed by law, the measure must also be sufficiently 'foreseeable'. We address this in the following paragraphs.

10.110 We consider that the measure is prescribed by law because it is set out with sufficient clarity and precision to enable providers to understand both the nature of the technology contemplated and how it should be applied in practice. In particular, the measure clearly identifies the use of hash matching technologies for the detection of terrorism content, the contexts in which such tools are expected to be deployed, and the key parameters relevant to their operation, including the use of appropriate hash datasets, appropriate matching thresholds, and supporting governance processes. Taken together, these elements ensure that the measure is accessible and foreseeable in its application, enabling providers to determine with reasonable certainty the steps required to comply, and thereby reducing the risk of arbitrary or inconsistent interference with lawful content and users' freedom of expression.

10.111 It is not appropriate or desirable to be prescriptive about the balance between precision and recall that providers should adopt. That is ultimately a judgement the providers have to take when considering this measure in the specific context of their own services. We have set out why it is important for a balance to be struck by providers when considering the impact on users' freedom of expression, in particular. We have provided further guidance on the appropriate proportion of human review in the draft Guidance on appropriate proportion of human review for hash matching, published alongside this statement.

10.112 In relation to the proportion of detected content subjected to human review, as set out in annex 4.3: Human review for proactive technology, we have amended the factors that providers should take into account when determining this. We have also set out why we consider it is not possible to be prescriptive about the proportion of detected content that should be subject to human review, and do not consider that this means that the measure is insufficiently foreseeable. We have also given some further guidance on the matters providers should consider in determining the appropriate proportion of human review.

10.113 In practice, service providers have legal incentives that should encourage them to strike an appropriate balance in their implementation of the measure. Service providers have political and reputational incentives to minimise the amount of terrorism on their platform. They also have incentives to limit the costs of moderation (including human review) and the handling of user complaints where content is wrongly taken down. We anticipate that providers will continuously adjust the technical configuration of their hash matching system in light of their experience.

Necessary for a legitimate aim

10.114 Any interference with the right to freedom of expression must pursue one or more legitimate aims.[754] In considering the measure's impact on individuals' human rights, our starting point is to recognise that there is a substantial public interest in reducing the online prevalence and dissemination of terrorism content. That public interest relates to the prevention and deterrence of terrorist crimes and protection of national security, the protection of morals of those who may be drawn into terrorism through exposure to the

[754] National security, territorial integrity or public safety; the prevention of disorder or crime; the protection of health or morals; the protection of the reputation or rights of others; preventing the disclosure of information received in confidence; and maintaining the authority and impartiality of the judiciary.

content and the protection of the rights and safety of those who may be the victims of terrorist crimes. The consequences of a successful terrorist attack can include many deaths.

10.115 We explained at paragraphs 10.58 – 10.59 why reactive complaints handling is not sufficient to address the harm caused: we are not aware of other effective ways to protect users from terrorism content besides using proactive technology to detect relevant content quickly. We also concluded that perceptual hash matching is more effective than cryptographic matching at protecting users from harm, notwithstanding the greater risk of false positives. It is against this background that we consider the proportionality of the measure's interference with freedom of expression.

Proportionality

10.116 As set out in paragraph 10.102 and in the relevant chapter in annex 1: Further stakeholder responses, we received several stakeholder comments expressing concern about the proportionality of the measure's interference with freedom of expression.

10.117 While we recognise these concerns, we remain of the view that the measure is proportionate to address the serious risk to individuals and society posed by terrorism content. As set out the section titled 'Effectiveness and benefits', there is evidence that terrorism content online continues to present a significant and evolving risk of harm, including by facilitating the spread of propaganda, supporting radicalisation pathways, and enabling the coordination and encouragement of terrorist activity, often at speed and scale across services. In light of this pressing need, we consider that the use of targeted technological measures, such as hash matching for known terrorism content, is proportionate: the measure is directed at a clearly defined category of serious illegal content and is limited to detecting material that has already been identified and verified (for example through trusted hash datasets), rather than requiring general or indiscriminate monitoring.

10.118 While there may be an interference with freedom of expression, it arises where there are mistakes, not as a matter of any particular kind of lawful content being systematically targeted. Incorrect detection of terrorism content does not lead to criminal or civil sanctions on those whose content is taken down. The content concerned may be taken down on one service, but this will not necessarily be the case on other services. Finally, the measure is implemented by service providers, and not by the state: it is not a tool through which to limit political speech.

10.119 In addition, the design of the measure, and the wider framework of law, includes a number of safeguards which protect users' freedom of expression. We have highlighted the main safeguards below before turning to our overall assessment of proportionality.

Databases

10.120 The risks to freedom of expression posed by the measure depend in part on the database used for hash matching. However, we have included a number of safeguards in the measure to secure that the database contains appropriate material. We have provided that the set of hashes must be sourced from a person with appropriate expertise in identifying terrorism material and that it must be secured against interference. The measure sets out that a service provider should take reasonable steps to secure the removal of a hash from the hash database where the provider has determined that the hash is not capable of amounting to terrorism content.

10.121 We have acknowledged in the statement that there are risks that hash matching may disproportionately affect minority groups. As noted at paragraph 10.72 we have not identified significant risks of bias arising from hash matching technology itself, given its deterministic nature. Bias could arise in the composition of the database. However, we note that historically in the UK, the list of proscribed organisations has included a large number of Islamist terrorist groups, and far fewer groups of other ideologies. To that extent, it is to be expected that an appropriate set of hashes would include more content from organisations of this nature than from other organisations. However, we have provided that an appropriate set of hashes must not plainly discriminate, in relation to identifying terrorism content, on the basis of protected characteristics and that it should include an appropriate range of terrorism content.

10.122 We have considered the Inclusion Policy of TAT and the way in which content is classified and are satisfied that TAT can be used in a way which focuses on content which amounts to illegal content in the UK. We do not consider that the Inclusion Policy discloses a risk of direct bias against any minority groups or of undue indirect bias.

Precision and recall configuration

10.123 Our measure aims to mitigate against the risk of false positives by recommending that providers strike an appropriate balance between precision and recall in configuring their technology. We expect service providers to take into account certain factors in striking this balance, including the risk of harm from terrorism content, the proportion of content which is not terrorism content detected as a match by the technology, and the effectiveness of the systems and processes used by the service provider to address this content. These principles aim to help ensure the accuracy and effectiveness of the technology and should therefore limit the risk of detection of content that is not terrorism content. We are also recommending that providers review the balance between precision and recall periodically which helps to ensure that providers continue to refine their approach in light of experience.

Review of detected content

10.124 Potentially illegal content detected through this measure is not automatically taken down. To safeguard against the potential risks of legal content being hashed and incorrectly classified, we have recommended that matches be moderated unless the provider has a high degree of assurance that the content is illegal content.[755]

10.125 The ICJG should assist providers when making judgements about whether content is terrorism content.

10.126 We have also recommended that certain providers have internal content policies;[756] have and monitor their performance against performance targets relating to accuracy; resource their content moderation function to give effect to these policies and targets;[757] and provide training to content moderators.[758] We expect that these will help to maintain the

[755] Technically, the review stage (content moderation) begins as soon as the detection takes place, because any subsequent decision that the degree of assurance is sufficiently high to take the content down is, itself, content moderation. However, it is likely that in most cases providers will need to deploy other layers of automated or human review before it is possible to take a decision on the content.
[756] Illegal Content Codes, Measure ICU C3.
[757] Illegal Content Codes, Measures ICU C4-C6.
[758] Illegal Content Codes, Measures IC7-C8.

quality of decision making so as to reduce impacts on lawful content. We also consider that they will help to prevent bias emerging and allow providers to identify it if it did. These measures only apply to large or multi-risk services.[759] We set out in our December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement) why we consider this to be proportionate. Although we consider it likely to be rare, we recognise that it is theoretically possible for a service to be at high risk of terrorism and not of any other illegal harm.

10.127 All regulated user-to-user service providers are subject to measure ICU C1 which requires them to have systems and processes designed to review and assess content the provider has reason to suspect may be illegal content. A provider may need to take external legal advice on matches in order to comply with this measure. Therefore, even where a service is not multi-risk, measure ICU C1 acts as a safeguard. We also consider that, where a service is not large or multi-risk, the impact of incorrect content moderation decisions is overall likely to be lower, although we acknowledge that the impact may nevertheless be significant for some users.

10.128 For quality assurance purposes, we are recommending that human moderators review an appropriate proportion of all detected content.

10.129 We consider that the human review factors will help mitigate the risks arising from the fact that many of the Schedule 5 terrorism offences require context in order to determine whether content is terrorism content or not. In addition, the recommendations in relation to human review will help mitigate risks arising where confidence in a match is lower and risks arising from any lack of quality assurance of the configuration of the technology. For example, where there is greater uncertainty as to the performance of the technology (or other technologies used as part of content moderation), the likelihood of false positives increases, and a higher proportion of human review will be appropriate. This helps to reduce the risk of providers incorrectly taking down content that is not terrorism content.

Appeals

10.130 An important safeguard in protecting lawful content from being removed is the appeal process.[760] In circumstances where content has been taken down, a user can appeal to have their content reinstated. Appeals provide a mechanism for individuals to inform the service provider that the content was removed or taken down incorrectly. This safeguards the speech (and potentially commercial interests) of such individuals.

10.131 The availability of user appeals further supports the proportionality of this measure. Where content is affected following detection through hash matching, providers are expected to offer accessible mechanisms for users to challenge moderation decisions, enabling errors to be identified and corrected. This provides an important safeguard against undue interference with lawful expression and helps ensure that decisions are subject to appropriate oversight and review.

10.132 We acknowledge that there will be an interference with freedom of expression while content remains unavailable pending any appeal. However, as set out in paragraphs 10.103 – 10.108 we do not think it is possible to fully mitigate this risk without significantly undermining the benefit of the measure.

---

[759] Illegal Content Codes, p.9.
[760] Illegal Content Codes, Measures ICU D8-D10.

Data protection law as a freedom of expression safeguard

10.133 Nothing in our Codes of Practice or guidance, insofar as it relates to the processing of personal data, either supersedes or derogates from the requirements of data protection legislation. These include the principle of accuracy[761] when processing personal data (e.g., the creation of data indicating that it is reasonable to infer that a user has committed a terrorism offence). Further, the UK General Data Protection Regulation (UK GDPR) restricts the circumstances in which decisions can be made based solely on automated processing of personal data, where the decision has legal or similarly significant effects for the relevant individual.

10.134 Services must comply with data protection law in implementing this measure and should have regard to relevant guidance issued by the ICO.[762] We therefore consider that service providers comply with their data protection obligations, this will support accurate, effective and free from bias decision-making about moderation of user-generated content and therefore safeguard users' rights to freedom of expression in respect of that content.

Overall assessment of proportionality

10.135 When considering the proportionality of the measure, our starting point is that the viewing and dissemination of terrorism content is a serious harm and reducing its dissemination has clear benefits for the public. Our recommendations for the review of content, as well as appeals, provide the most important safeguards in this regard.

10.136 Overall, we consider that the measure does not give rise to a disproportionate interference with freedom of expression, taking into account its targeted scope, the safeguards in place, and the importance of addressing the harms associated with terrorism content.

Privacy

10.137 We received stakeholder responses to the June 2025 Consultation relating generally to privacy and data protection concerns. For example, an individual respondent warned that the act of reviewing content flagged as potentially illegal has serious problems related to user privacy.763

10.138 The ICO welcomed Ofcom's recognition of the need to comply with data protection law when using hash matching, and its reference to ICO's guidance on content moderation. It said personal data is likely to be processed during hash matching and database creation, and service providers must comply with data protection law.764

10.139 The ICO also noted that the use of hash matching technologies may, depending on the circumstances, engage provisions relating to automated decision-making under UK data protection law. It stated that service providers must consider whether restrictions relating to solely automated decisions apply and ensure compliance with any relevant legal requirements and safeguards.[765]

Is there an interference?

10.140 The measure recommends that service providers deploy hash matching technology to detect terrorism content communicated publicly, not content communicated privately.

---

[761] Information Commission's Office, Principle (d): Accuracy | ICO.
[762] Information Commission's Office, UK GDPR guidance and resources | ICO.
[763] Name Withheld 4 response to the June 2025 Consultation, p. 5.
[764] ICO response to the June 2025 consultation, p.15.
[765] ICO response to the June 2025 consultation, p.16.

Under the Act, Ofcom is unable to recommend in Codes of Practice the use of proactive technology to content communicated privately.766 Individuals posting public content have a lesser expectation of privacy, and therefore the extent of any intrusion into privacy is correspondingly reduced.767 However, we accept that it is possible for content communicated publicly to engage the right to privacy.

10.141 The measure recommends that providers use a suitable database for the purpose of hash matching. The extent to which the inclusion of hashes on a database interferes with privacy varies depending on the database. The database maintenance involves the processing of personal data, any interference with privacy is limited and incidental, arising only where necessary to ensure that terrorism content is effectively identified and its visibility reduced. In practice, such processing is minimal and directed towards protective purposes and therefore does not constitute a significant interference with users' privacy.

10.142 Hash matching involves the automated processing of content communicated publicly on a service. Service providers may also use automated technology to review detected content as part of their content moderation systems. The ICO commented that compliance with data protection law does not, in itself, guarantee that the privacy impact on users will be minimal.[768] We agree; nevertheless, we remain of the view that the hashing of users' content in and of itself involves only minimal intrusion with their privacy rights under the ECHR.[769] The processing is limited in nature, consisting of the conversion of content into a hash and its comparison against known hashes, without any immediate or determinative assessment of legality. The activity is carried out by private entities rather than the state and therefore does not amount to state surveillance. A match may give rise to an initial indication that content could relate to terrorism, but this is not conclusive and requires further assessment, while non-matched content is not implicitly deemed lawful. On this basis, any interference with privacy is limited in scope and preliminary in nature.

10.143 As a result of the measure, a greater volume of content may be subject to review within providers' moderation processes, including by human moderators, which entails a degree of additional intrusion into users' privacy. However, this intrusion is reduced by the fact that the content has already been communicated publicly, and the review is limited to assessing material that is already accessible on the service. Any interference with privacy is therefore constrained and arises in the context of targeted measures aimed at identifying potentially harmful content.

In accordance with the law

10.144 We have explained in paragraphs 10.109 – 10.113 why our measure satisfies the requirement that any interference be 'prescribed by law'. For the same reasons we consider that the measure is 'in accordance with law' for the purpose of Article 8.

---

766 Schedule 4 of the Act, paragraphs 13(4). We have set out guidance which is intended to assist services in deciding whether content is communicated 'publicly' or 'privately' for this purpose: Guidance on content communicated 'publicly' and 'privately' under the Online Safety Act.

767 The ICO also noted that compliance with data protection obligations is required regardless of whether content is communicated publicly or privately: ICO response to our June 2025 Consultation, p.12.

768 ICO response to the June 2025 Consultation, p15.

769 We note that there are specific data protection considerations for automated data processing. See (at the time of this Statement): ICO, 2026. Automated decision-making, including profiling.

Necessary for a legitimate aim and proportionate

10.145 The measure pursues the legitimate aim of protecting users, including children, from the harms associated with terrorism content online, including its role in facilitating radicalisation, disseminating propaganda, and contributing to serious real-world harms.

10.146 As set out in the chapter 8: Introduction to proactive technology, some degree of processing of user content is necessary to enable providers to identify and mitigate illegal content effectively; purely manual approaches would be significantly slower, less reliable, and potentially more intrusive. This is underscored by the evidence, outlined above, of the serious and growing harms associated with terrorism content online, including its role in facilitating radicalisation and dissemination at scale, which reinforces the necessity of targeted technological measures.

10.147 The measure also recommends that providers configure their technology to strike an appropriate balance between precision and recall, taking into account certain factors. We consider the recommendations in our measure in relation to precision and recall also support the accuracy and fairness principles of data protection law, and provides an important privacy safeguard against content being erroneously moderated.

10.148 To the extent that our measure recommends human review, this may be viewed as more privacy intrusive than automated review. However, human review is an important safeguard against taking down of lawful content; it therefore helps to protect the freedom of expression of users posting lawful content. It also mitigates the risk of inaccurate personal data being created when content is incorrectly determined to be illegal and supports the fairness principle in data protection law.[770] Overall, we consider that the interference posed by human review is proportionate and that there is no less intrusive way to achieve our policy objective.

Appeals

10.149 Users can appeal if they believe a provider has taken action against them or their content based on an incorrect illegal content judgement. This enables them to mitigate the risk of their lawful content being taken down in the future, and to rectify inaccurate personal data held by the service provider.

Data protection law as a privacy safeguard

10.150 As set out in paragraphs 10.133 – 10.134 in relation to freedom of expression, the UK GDPR has the effect of protecting privacy to the extent that an interference with privacy may result from a breach of the rules on the processing of personal data.

10.151 Services must comply with data protection law in implementing this measure and should have regard to relevant guidance issued by the ICO. In particular, the data protection

[770] ICO response to the June 2025 consultation, p.17. The ICO noted that Ofcom's recommendation that services should review an appropriate proportion of detected content is an important privacy and data protection safeguard to guard against content being erroneously detected as terrorism content and support the accuracy and fairness principles of data protection law.

principles of data minimisation,[771] accuracy[772] and fairness[773] will serve to protect privacy, and providers must have regard to them (and all other relevant principles) when processing personal data as part of the measure. Automated decision-making is only permitted where service providers have implemented certain safeguards for the data subject's rights, freedoms, and legitimate interests in accordance with the UK GDPR.[774]

Other safeguards

10.152 We expect service providers to have particular regard to the importance of protecting users' privacy (among other things) when complying with their safety duties.[775] Users can complain to service providers if they consider that the provider is not complying with this duty.[776] This is a further safeguard provided by the Act.

Overall assessment of proportionality

10.153 Overall, we consider that the measure does not give rise to a disproportionate interference with users' privacy, taking into account its limited scope, targeted application, and the safeguards in place.

Who the measure applies to

10.154 In our June 2025 Consultation we proposed that the measure would apply to providers of user-to-user services that enable user generated content in the form of photographs, videos or visual materials to be generated, shared or uploaded and are:

a) large services at medium or high risk of terrorism content; or
b) at high risk of terrorism content and:
    i) have more than 700,000 monthly UK active users; or
    ii) are a file-storage and file-sharing service.

10.155 We received a range of feedback from industry and an individual stakeholder who raised concerns about the proportionality of this measure due to wider concerns about the financial cost of implementing the measure, its impact on smaller services,[777] and access to effective third-party datasets.778 We discuss our proportionality assessment for the different groups of services in scope of this measure below.

10.156 Amending our measure to recommend that some services use a third-party dataset to develop their database of hashes means there may be additional costs of implementing this measure.779 We consider this change to be proportionate for providers of in-scope services as it substantially improves the effectiveness of hash matching where it applies. In addition, for some small services, TAT is able to provide access to its hash list API at a

[771] Information commission's office, Principle (c): Data minimisation | ICO (this guidance is currently under review and may be subject to change).
[772] Information commission's office, Principle (d): Accuracy | ICO (this guidance is currently under review and may be subject to change).
[773] Information commission's office, Principle (a): Lawfulness, fairness and transparency | ICO (this guidance is currently under review and may be subject to change).
[774] Information commission's office, Rights related to automated decision making including profiling | ICO
[775] Section 22 of the Act.
[776] Section 21(4)(b)(iii) of the Act.
[777] ACT – The App Association response to the June 2025 Consultation, p.7.
[778] [✂].
[779] See 'How to compile the hash database' in 'How this measure works' for further detail on the criteria a third-party provider would need to meet in order to be considered as having an appropriate set of hashes.

reduced cost or, in some cases, free of charge.[780] This will help mitigate the additional cost burden for some smaller in-scope services.

Inclusion of medium risk large user-to-user services

10.157 One respondent agreed with applying the measure to large services, high risk services and file-sharing/file-storage services.781

10.158 Large services have the greatest volumes of content and their large user bases means that there is a risk that a piece of terrorism content could be encountered by a significant number of users. Given the potential of harm to many users and because providers of large services typically have greater financial resources, we consider it proportionate to recommend this measure to providers of large services at medium or high risk of terrorism content.

10.159 We consider that these services will have the financial and technical capacity to implement hash matching for terrorism, either by developing their database using terrorism content identified through past decisions or by sourcing their database from a third-party hash list. Some of the largest services already use a form of hash matching for terrorism.[782]

Inclusion of high risk user-to-user services with more than 700,000 active UK users

10.160 One respondent raised concerns about the proportionality of this measure. They said that reliable detection tools and datasets for terrorist content do not currently exist at scale, and that while hash matching may be suited to some large-scale platforms, smaller or lower-risk services may lack the technical or financial capacity to implement it effectively.783

10.161 We maintain our recommendation that it is beneficial for high-risk service providers with more than 700,000 active UK users to implement hash matching for terrorism content. The severity of terrorism content means there is a potential for serious harm to both users and wider society. We therefore consider that certain services that are high risk for terrorist content should do more to mitigate this type of content even if they are not large. While costs may increase with the amount of such content on a service, we also expect the benefits of hash matching to increase correspondingly. There is also a risk that terrorist actors may migrate to smaller services if hash matching technology is adopted predominantly by larger services.[784]

10.162 By restricting the scope to those services with more than 700,000 users that are high-risk, we target services which present the greatest potential harm while limiting the impact on services which demonstrate lower risk. We consider the threshold of 700,000 users to include services which are still able to meet the costs of this measure. This threshold also aligns with our measures on hash matching for CSAM and intimate image abuse so providers of services in scope of different hash matching measures may be able to take advantage of cost synergies.[785] We therefore consider it proportionate to recommend this measure to providers of services that have more than 700,000 monthly active users and that are high risk for terrorism content.

---

[780] Tech Against Terrorism, Hash List Access Tiers.
[781] [✂].
[782] Global Internet Forum to Counter Terrorism (GIFCT), Membership.
[783] [✂].
[784] Romaniuk, S.N., Fabe, A.P.H. and Nandy, D. 2023, Terrorist Platform Migration: The Move to Smaller, Less Regulated Online Spaces. GNET Insights, 23 June.
[785] Ofcom (2026), May 2026 Statement, Detecting Intimate Image Abuse.

Inclusion of high risk file-sharing and file-storage services

10.163 All file-sharing and file-storage services at a high risk for terrorism content should employ hash matching to identify this content, regardless of size. We did not receive feedback that disagreed with the inclusion of file-sharing and file-storage services as a specific service type. However, one stakeholder raised concerns that the cost of doing this would be disproportionate for small services.[786]

10.164 We recognise that the financial impact the measure could have on providers of small file-sharing services is potentially significant. However, our analysis suggests that file-sharing services, including small ones, tend to pose an especially high risk in relation to terrorism content. As outlined in our June 2025 Consultation, research conducted by TAT found that between November 2020 and January 2023, file-sharing services represented more than half of services where they identified terrorist content.[787] Our Illegal Harms Register of Risks (Illegal Harms Register) also outlined how file-sharing and file-storage services are critical to the storage of the majority of terrorism content.[788] Many of these services will have fewer than 700,000 monthly active UK users. We also received feedback from a stakeholder which agreed with the inclusion of file-sharing services.[789] Research from GIFCT provided additional insights on the risk of harm that file-sharing services pose in hosting content that is removed from other services.[790]

10.165 Given the severity of the impact that the dissemination of terrorism content has, we conclude that it is proportionate to include all file-storage or file-sharing services which are at high risk of terrorism content in scope of the measure regardless of size, despite the potentially significant cost impact.

10.166 Most small file-sharing services would need to develop their database of hashes using a third-party hash list.

10.167 There is a risk that some small file-sharing services may choose not to operate in the UK as a result of this measure. We consider this a proportionate outcome given the potential harm posed by such services operating without hash matching to detect terrorism content.

10.168 Although high-risk file sharing and file storage services are in scope of this measure, the measure only recommends that service providers deploy hash matching technology to detect terrorism content communicated publicly, not content communicated privately.[791] We discuss this in more detail in chapter 8: Introduction to proactive technology.

Conclusion

10.169 Taking account of the severity of the risk of harm to users and other individuals associated with the dissemination of terrorism content online, we have decided to include the measure on the use of perceptual hash matching for terrorism content in our Illegal Content Codes.

---

[786] ACT – The App Association response to the June 2025 Consultation, pp.6-7.
[787] Tech Against Terrorism (2023), Patterns of Online Terrorist Exploitation.
[788] Ofcom (2024), Illegal Harms Register of Risks (Illegal Harms Register), p.40.
[789] [✂].
[790] GIFCT, 2026. GIFCT Partners with UK Home Office on File Sharing Report. GIFCT News, 13 August.
[791] Schedule 4 of the Act, paragraphs 13(4). We have set out guidance which is intended to assist services in deciding whether content is communicated 'publicly' or 'privately' for this purpose here: Guidance on content communicated 'publicly' and 'privately' under the Online Safety Act.

10.170 This measure applies to providers of user-to-user services that enable regulated user-generated content in the form of photographs, videos or visual materials to be generated, uploaded, or shared and:

a) are large services that are medium or high risk of terrorism content; or
b) are at high risk of terrorism content, and:

i) have more than 700,000 active UK users; or
ii) are file-sharing or file-storage services.

10.171 We recognise that the measure imposes costs on service providers. However, we consider these costs to be proportionate in light of the seriousness of the harms the measure seeks to address and the benefits of reducing the dissemination of terrorism content online. The availability of a third-party dataset, together with the flexibility afforded to providers in implementation, helps to reduce the burden on services while maintaining the measure's effectiveness.

10.172 We also acknowledge that the use of hash matching may engage users' rights to freedom of expression and privacy, particularly where lawful content is incorrectly identified. For this reason, the measure includes a range of safeguards, including recommendations relating to dataset quality, the appropriate configuration of matching technology, human review, appeals, and compliance with data protection law. These safeguards are intended to minimise the risk of erroneous moderation and undue interference with users' rights.

10.173 Taking these factors together, we consider that the measure strikes an appropriate balance between reducing the significant harms associated with terrorism content and protecting the rights and interests of users and service providers. We therefore conclude that the measure is proportionate when assessed in the round. The full text of the measure can be found in our Terrorism Code of Practice for user-to-user services and is referred to as ICU C13.

Extending perceptual hash matching for child sexual abuse material (CSAM)

10.174 In the December 2024 Statement we included a measure (ICU C9) in our Illegal Content Codes setting out that providers of the following types of service should use perceptual hash matching[792] to detect image-based child sexual abuse material (CSAM):[793]

- Large services that are at medium or high risk of image-based CSAM.
- Services that are at high risk of image-based CSAM and (a) have more than 700,000 monthly active UK users or (b) are file-storage and file-sharing services.

10.175 In the June 2025 Consultation, we proposed widening the scope of the measure so that perceptual hash matching also be deployed by all services (regardless of size) whose principal purpose is hosting or disseminating regulated pornographic content and who are at high risk of image-based CSAM.

[792] Our definition for "perceptual hash matching" can be found in chapter 8: Introduction to proactive technology. By "hash matching", we mean "perceptual hash matching" and we use both terms interchangeably throughout this chapter.
[793] Illegal Content Codes, Measure ICU C9.

Our decision

10.176 We have decided to proceed with the extension of the measure to cover all services (regardless of size) whose principal purpose is hosting or disseminating regulated pornographic content and who are at high risk of image-based CSAM, as proposed in our June 2025 Consultation.

10.177 Having considered responses to the consultation, we have also decided to make some changes to the principles to which a provider should have regard in determining the "appropriate proportion" of detected content that should be subject to human review. Our objective in making this change is to be clearer about what these principles are for, and to align these with our measures covering hash matching for terrorism and intimate image abuse (IIA). Our reasoning is explained in full in annex 4.3: Human review for proactive technology. We have also provided separate guidance for providers on this matter in our 'Draft guidance on appropriate proportion of human review for hash matching', published alongside this statement.

10.178 The full text of the updated measure can be found in our Illegal Content user-to-user Codes of Practice and is referred to as ICU C9. The measure is part of our package of measures to target child sexual exploitation and abuse (CSEA).

How this extension works

10.179 Full details of how the hash matching measure works are available in the December 2024 Statement.[794] This includes the recommendations we set out for providers to ensure that the set of hashes that providers are using is appropriate for the measure. It also includes several safeguards designed to minimise false positives and ensure that they are managed by providers in a rights-preserving way, as well as recommendations for human moderation as part of this measure.[795] [796]

Who it applies to

10.180 As a result of our decision, this measure will now also apply to services at high-risk of image-based CSAM whose primary purpose is hosting or disseminating regulated pornography content.

---

[794] Ofcom (2024), December 2024 Statement, Vol.2, chapter 4: Automated content moderation.

[795] We explain that the design of the design of measure ICU C9 includes a number of safeguards to protect users' freedom of expression directed at the set of hashes used and the technology's configuration. This includes, for example, arrangements put in place to ensure that child sexual abuse material (CSAM) is correctly identified before hashes of that material are assed to the set of hashes used, and that the provider should review any cases where material is suspected to have been incorrectly identified, alongside other examples. A full discussion of freedom of expression impacts and safeguards are set out in the December 2024 Statement, Volume 2: Service Design and User Choice, paragraphs 4.148-4.162.

[796] We also specify in the December 2024 Statement, Volume 2: Service Design and User Choice, chapter 4, paragraphs 4.154-4.155, that measure ICU C9 sets out other processes designed to further reduce the impact of false positives detected by the technology. This includes, for example, that service providers should put in place a policy of review for detected content which secures that human moderators review an appropriate proportion of content detected as CSAM; and that our other Codes measures act as safeguards for users' freedom of expression.

10.181 We define 'regulated pornography content' as pornographic content other than content of a type described in section 61(6) of the Act.797 This definition includes pornography that is not user-generated content. As such, a Part 5 service which also hosts a small amount of user-generated content is caught by the measure, no matter how small the proportion of content, and will be expected to hash match.[798] However, "illegal content" is defined as user-generated content,[799] so they would only need to apply the hash matching measure to the user-generated content they host.

Our reasoning

10.182 This section outlines the evidence demonstrating that extending the measure will have significant benefits for users. This includes an explanation of the amendments to the factors that providers should take into account when deciding what proportion of detected content should be subject to human review. Finally, it sets out the impacts and costs of extending the measure, including discussing the rights impact.

Effectiveness and benefits

10.183 In the December 2024 Statement, we set out that removing known CSAM online will deliver important and wide-reaching benefits.[800] The use of perceptual hash matching is a well-established industry best practice and has resulted in the detection and removal of millions of items of CSAM.[801] [802] Hash matching is therefore an important tool for detecting and removing CSAM. We are now recommending a new measure to use proactive technology to detect unknown image-based. We see these measures as complementary and explain our reasoning in chapter 9: Deployment of proactive technology to detect image-based CSAM.

10.184 Pornography services are known to be a high risk area for CSAM, being one of the most common types of online service where CSAM is encountered on the clear web.[803] [804] Pornography services are often used by offenders in seeking and searching for CSAM.[805]

---

[797] Section 61(6) excludes from the definition of 'regulated pornographic content', content which consists only of text, or consists only of text accompanied by identifying content which consists only of text, other identifying content which is not itself pornographic content, a GIF which is not itself pornographic content, an emoji or other symbol, or any combination of these.

[798] The measures only apply to regulated user-to-user services. A service may be a regulated user-to-user service even if the amount of user-generated content on the service is small (see section 3(2)(b) of the Act). In the case of a regulated user-to-user service whose primary purpose was the hosting of regulated pornographic content, the Code measure would apply to the service regardless of whether the pornographic content concerned was primarily provider-generated or user-generated content.

[799] See section 59(14) of the Act.

[800] Ofcom (2024), December 2024 Statement, Vol.2, chapter 4: Automated content moderation.

[801] Ofcom (2024), December 2024 Statement, Vol.2, chapter 4: Automated content moderation; Ofcom (2024), Illegal Harms Register of Risks, chapter 2B: Child Sexual Abuse Material (CSAM).

[802] Ofcom (2024), December 2024 Statement, Annex 5: Assumptions on Costs and further analysis on costs and benefits, p.12.

[803] For example, Protect Children Finland found that 31.7% of offenders surveyed had encountered CSAM on a pornography service, making it the most common type of online service identified in the survey for where CSAM was encountered on the clear web within that survey. Protect Children (2024), Tech Platforms Used by Online Child Sexual Abuse Offenders, p.29.

[804] We define the 'clear web' as publicly available internet used for everyday activities.

[805] Ofcom commissioned Protect Children, a global expert in research concerning child sexual exploitation and abuse (CSEA), to carry out an anonymous survey among perpetrators who have used known keyword terms to search for CSAM on the dark web. Participants took part voluntarily and were not renumerated. The report finds that the majority of respondents reported viewing material involving children aged 11 to 17. 50% of

Protect Children's March 2026 CSAM Perpetrator Research Report, commissioned by Ofcom, found that 22% of respondents used pornography services to search for CSAM.[806] Moreover, it was common for users (who were often children at the time) to be exposed to CSAM unintentionally, including on pornography services.[807]

10.185 Numerous pornography services reported to us to be hosting CSAM fall under the current 700,000 user threshold, with the actual number of such services hosting CSAM likely to be higher.[808] Therefore, expanding the scope of the hashing measure to include all high risk pornography services regardless of size will materially increase the amount of CSAM that is detected and removed, thereby delivering significant benefits. A number of responses to the June 2025 Consultation reinforce this conclusion. For example, a number of respondents expressed support for the extension, and several stakeholders corroborated our assessment that hash matching is effective at addressing CSAM and its associated harm.[809] While some respondents disagreed with our assessment, none provided any convincing evidence to refute it.

10.186 Our view that it is appropriate to expand the scope of the hashing measure is reinforced by our current understanding of the ecosystem of hash databases. We set our original thresholds as we were concerned existing database providers would not have the capacity to manage the influx. We are now confident that the ecosystem has sufficient capacity for us to expand the scope as proposed. We have observed that hash database providers have been able to accommodate the increased demand for their services following the

---

respondents view material depicting 11–14-year-olds and 51% view material depicting 15-17-year-olds. Protect Children (2026) Tell Me More About Tech CSAM Perpetrator Research Report, p.16.

[806] Protect Children (2026), Tell Me More About Tech CSAM Perpetrator Research Report, p.20.

[807] Protect Children (2026), Tell Me More About Tech CSAM Perpetrator Research Report, p.15.

[808] As we explain in our June 2025 Consultation, confidential evidence seen by Ofcom of pornography services found hosting CSAM showed that there were up to 247 pornography services found hosting CSAM in 2024, with the actual number of such services likely to be higher. The majority of these services were found to host post-pubescent CSAM. Email correspondence with IWF, 27 March 2025. See Ofcom (2025), Additional Safety Measures Consultation, paragraph 13.6.

[809] A member of the Philippine Survivor Network response to the June 2025 Consultation, pp.19-20; Barnardo's response to the June 2025 Consultation, p.13; British and Irish Law Education and Technology Association (BILETA) response to the June 2025 Consultation, p.19; Canadian Centre for Child Protection (C3P) response to the June 2025 Consultation, p.14; CCO response to the June 2025 Consultation, pp.17-18; Commissioner Designate for Victims Of Crime Northern Ireland (CVOCNI) response to the June 2025 Consultation, p.5; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.11; Domestic Abuse Commissioner response to the June 2025 Consultation, p.1; Evangelical Alliance response to the June 2025 Consultation, p.8; Free Speech Union response to the June 2025 Consultation, p.5; [✂]; ICO response to the June 2025 Consultation, p.18; Internet Watch Foundation (IWF) response to the June 2025 Consultation, p.30; Lucy Faithfull Foundation response to the June 2025 Consultation, p.6; [✂]; National Center for Missing & Exploited Children (NCMEC) response to the June 2025 Consultation, p.6; [✂]; The National Society for the Prevention of Cruelty to Children (NSPCC) response to the June 2025 Consultation, pp.28-29; Online Dating and Discovery Association (ODDA) response to the June 2025 Consultation, p.5; Online Safety Network response to the June 2025 Consultation, p.20; [✂]; Refuge response to the June 2025 Consultation, p.13; [✂]; [✂]; We Are Survivors response to the June 2025 Consultation, p.5.

introduction of our Illegal Content Codes.[810] Stakeholder responses to the June 2025 Consultation have reinforced this belief.[811]

Impacts and costs

Direct costs

10.187 In our December 2024 Statement we set out the estimated costs for the implementation of perceptual hash matching for CSAM.[812] We maintain our view that the main costs are likely to consist of:

a) one-off costs related to building a hash matching system;
b) annual costs relating to maintaining a hash matching system;
c) costs of software, hardware, and data; and
d) costs of reviewing matches, moderating content, and reporting CSAM to external bodies.

10.188 In Table 10.2 we present the current estimates for these categories of costs. These estimates are based on the same methodology and labour assumptions used for this measure in the December 2024 Statement but have been updated to reflect the latest salary data from the Office for National Statistics (ONS).[813]

Table 10.2: Estimated costs to service providers of implementing hash matching for CSAM

Component	Estimated cost
Build cost (one-off)	£22,000 to £387,000
Maintenance costs (annual)	£5,000 to £97,000
Software, hardware, and data (annual)	From £1,000 for a small service
Human moderation (annual)	£22,000 to £67,000[814]

10.189 The costs to service providers represent a significant range and will vary based on the type of service, its complexity and size of its user base.

---

[810] There are at least five known providers of hash matching databases for CSAM. We have not seen any signs of these struggling to cope following the introduction of our Illegal Content Codes, and many of the organisations responsible for maintaining these databases have submitted consultation responses in support of extending the measure. Furthermore, the IWF has recently launched Image Intercept, a free tool to help smaller, eligible companies and startups proactively detect and remove known CSAM material on their services.

[811] BILETA response to the June 2025 Consultation, p.19; IWF response to the June 2025 Consultation, pp.33-36; Lucy Faithfull Foundation response to the June 2025 Consultation, p.6; NCMEC response to the June 2025 Consultation, p.6; [✂].

[812] Ofcom (2024), December 2024 Statement, Vol.2, chapter 4: Automated content moderation.

[813] We assume the implementation (build cost) would require 2 to 18 months of software engineering time, matched with equal time from professional occupation staff. Our standard assumption is that annual maintenance costs are 25% of the build cost. See Annex 5 for further details on our economic assumptions and analysis.

[814] Our methodology for calculating the cost of moderation is explained in Annex 5 of our December 2024 Statement. Our estimate of £22,000 to £67,000 is what we may expect a high-risk service with 700,000 users to incur. Smaller and less risky services would have moderation costs lower than this while large services would have higher costs than the range we have estimated.

10.190 We received support for our cost assessment from a range of stakeholders.[815] Some industry and individual stakeholders disagreed with our impact assessment and our assessment of costs.[816] These generally covered two areas: costs that were not considered in our assessment, and the overall magnitude of our cost estimates.

10.191 These responses were not related specifically to the extension of this measure to high-risk pornography services and were previously addressed in the December 2024 Statement. Our position on these points is set out in the relevant chapter in annex 1: Further stakeholder responses.

10.192 Having considered the feedback relating to our impact assessment, we maintain our approach to estimating costs. We consider the estimates in Table 10.2 to accurately reflect the range of costs incurred by providers which are now in scope of this measure.

Rights impact

Freedom of expression and privacy

10.193 We assessed the freedom of expression impacts of this measure in paragraphs 4.148-4.162 of the December 2024 Statement and the privacy impacts of this measure in paragraphs 4.163-4.177 of the December 2024 Statement. As set out in chapter 9: Deployment of proactive technology to detect image-based CSAM, we consider that the use of hash matching remains necessary and proportionate notwithstanding that we now also recommend the use of proactive technology. As such, we consider the reasoning in our December 2024 Statement to continue to apply, and in particular to apply in relation to the extension of the measure to more services.

10.194 We received some stakeholder responses which were not related specifically to the extension of this measure to high-risk pornography services. We consider these in the relevant chapter in annex 1: Further stakeholder responses.

Conclusion on the extension

10.195 The available evidence, as set out in the section titled 'Effectiveness and benefits' has shown that pornography services are one of the most common types of online service where CSAM is encountered on the clear web.

10.196 The evidence also suggests that smaller pornography services can play a significant role in the dissemination of CSAM. As a result, it is our view that all services whose principal purpose is the hosting or dissemination of regulated pornographic content who are at high risk of image-based CSAM, regardless of size, should implement the measure.

10.197 As set out in the 'Impacts and costs' section, the costs incurred as a result of hash matching could vary substantially from service to service. It is possible that the costs could prove to be a significant burden for some smaller pornography services. Nonetheless, the harm caused by CSAM is incredibly severe. In this context, we consider the costs to be proportionate.

---

[815] A member of the Philippine Survivor Network response to the June 2025 Consultation, pp.19-20; BILETA response to the June 2025 Consultation, p.19; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, p.10; [✂]; Name Withheld 4 response to the June 2025 Consultation, p.7.
[816] [✂]; [✂]; Name Withheld 9 response to the June 2025 Consultation, p.9; Pinterest response to the June 2025 Consultation, p.4; [✂].

10.198 For these reasons, we have concluded that it is appropriate and proportionate to extend the scope of application of our CSAM hash matching measure in the Illegal Content Codes so that it applies to all services (regardless of size) whose primary purpose is hosting or disseminating regulated pornographic content and who are at high risk of image-based CSAM.

10.199 This extension will be included in the Illegal Content Codes of Practice as an addition to ICU C9.

11. Recommender systems

What is this chapter about?

Recommender systems use an algorithm to curate personal content feeds and are often the primary method for users discovering new content on user-to-user services. However, recommender systems can also expose users to harmful and illegal content. If illegal content is not identified through moderation systems quickly, a recommender system may lead to the rapid spread of this content.

In our Protection of Children Code, we said that providers of in scope services should ensure that recommender systems exclude potential primary priority content from the feeds of any user who is not identified to be an adult using highly effective age assurance.

In this chapter, we set out our decision to proceed with measures to exclude content indicated to be potentially relevant illegal content[817] from children's recommender feeds.

What decisions have we made?

In response to stakeholder feedback, we have amended our measure to target children's recommender feeds rather than all feeds. We have also decided not to apply the measure to foreign interference offences.

Number in Codes	Recommended measure	Who should implement this
ICU E3	Content recommender systems: excluding potential relevant illegal content.	Services likely to be accessed by children that: a) are at medium or high risk of suicide and self-harm; and b) where a child-accessible part of the service has a content recommender system.
ICU E2	Content recommender systems: excluding potential relevant illegal content.	Services likely to be accessed by children that a) are at medium or high risk of hate and/or b) are medium or high risk of terrorism and c) where a child-accessible part of the service has a content recommender system.

Why have we made these decisions?

This measure will reduce the risk of illegal content being recommended to children. By doing so, it should deliver important benefits to children.

[817] Throughout the chapter, we have shortened to "potentially relevant illegal content" to "potentially illegal content."

Introduction

11.1 Content recommender systems are algorithmic systems used to curate personalised content feeds on user-to-user services and can be a primary method for users to discover other users' content. Recommender systems match users with content that is likely to interest them, encouraging greater engagement. Content is recommended based on factors such as popularity of content, characteristics of the content, characteristics of a user, user behaviour, preferences, and/or predicted engagement. Content presented to the user by the recommender systems includes regulated user-generated content such as images, videos, posts, and livestreams.

11.2 Our definition of 'content recommender systems' does not include the following:

a) recommender systems that suggest content to users in direct response to a search query,
b) network recommender systems, which suggest other users to follow or groups to join (such as network expansion prompts),
c) product recommender systems that are used exclusively for the purpose of recommending goods and services, or
d) content recommender systems that only recommend a users' own content from their own private inventory where this is not capable of being encountered by another user.[818]

11.3 Content recommender systems can help users discover new and engaging content. However, they can also expose users to harmful content, including illegal content they have not actively sought out but may be recommended based on wider inferences about the user's interests.[819] If this illegal content is not quickly detected and taken down by existing moderation systems, recommender systems may spread that content further and expose more users to it before it is taken down. Being served more of this harmful content can cause cumulative harm.

11.4 Over time, these systems can also lead users to experience a 'filter bubble' (increased exposure over time to thematically homogenous content or content with a unified theme) or 'rabbit hole' (content of increasing thematic intensity) effect.[820] [821] [822]

11.5 We also consider the risk of harm to users may increase through the adversarial usage or 'gaming' of the recommender systems. Some users may create or upload inflammatory content or content with a high shock value to exploit recommender systems and maximise

[818] We understand that photo sharing and storage services may use private or closed content recommender systems to curate personalised 'albums' based on specific themes which would be limited to a user's private content. Such recommendations of a user's own private content would not meet the definition of user generated content under the Act if they were not capable of being encountered by another user.
[819] Ofcom (2025), Children's Register of Risks (Children's Register), p.50-52, 79-81, 107-109, 142-144, 190-191, 213-214, 227-228, 247-248; Ofcom (2025), Illegal Harms Register of Risks (Illegal Harms Register), p.50, 119-120, 319-320, 344.
[820] RUSI (Reed, A., Whittaker, J., Votta, F. and Looney, S.), 2019. Radical Filter Bubbles: Social Media Personalisation Algorithms and Extremist Content.
[821] A rabbit hole is where a user experiences a sequence of content recommendations that are increasingly more extreme in nature in relation to a particular subject matter or harm. This may occur as a result of users regularly engaging with content of a specific nature.
[822] 5Rights Foundation response to the June 2025 Additional Safety Measures Consultation (June 2025 Consultation), pp.10-11.

engagement. As explained in the Illegal Harms Register of Risks (Illegal Harms Register), this is likely to occur in online communities that share 'edgy' humour or 'trashposting' and may increase the recommendation of hate content.[823] [824]

11.6 To address the risks from recommender systems, we have already recommended the following measures in our Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes) and Protection of Children Code of Practice for user-to-user services (Protection of Children Code). [825] [826]

Illegal Content Codes

a) Measure ICU E1: Providers of services in scope should collect safety metrics during on-platform testing of recommender systems. The safety metrics should enable providers to understand whether a content recommender system design adjustment would increase the risk of users encountering illegal content.

Protection of Children Code

a) Measure PCU E1: Providers of services in scope should ensure content recommender systems are designed and operated so that content indicated potentially to be primary priority content on the basis of relevant available information is excluded from relevant users' content feeds (UK user of the service, other than any user determined to be an adult by the use of highly effective age assurance).[827]
b) Measure PCU E2: Providers of services in scope should ensure content recommender systems are designed and operated so that content indicated potentially to be either priority content or an identified kind of non-designated content on the basis of relevant available information is excluded from, or given a low degree of prominence in, relevant users' content feeds.[828] [829]
c) Measure PCU E3: Providers of services in scope should enable child users to give negative feedback on recommended content, and design and operate recommender systems to ensure that this feedback is taken into account for the purpose of reducing the likelihood of that user encountering similar content.

11.7 In our June 2025 Additional Safety Measures Consultation (June 2025 Consultation), we proposed an additional measure to reduce the likelihood of users encountering 'content indicated potentially to be relevant content' on their recommender feeds.

---

[823] Ofcom (2025), Illegal Harms Register, p.12-13.
[824] Munn, L., 2020. Angry by design: Toxic communication and technical architectures. Humanities and Social Sciences communications. 7 (53).
[825] Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes).
[826] Protection of Children Code of Practice for user-to-user services (Protection of Children Code).
[827] Primary priority content refers to pornographic content, and content which encourages, promotes, or provides instructions for suicide; self-harm; or an eating disorder or behaviours associated with an eating disorder.
[828] Priority content refers to content that is abusive on the basis of race, religion, sex, sexual orientation, disability or gender reassignment; content that incites hatred against people on the basis of race, religion, sex, sexual orientation, disability or gender reassignment; content that encourages, promotes, or provides instructions for serious violence against a person; bullying content; content which depicts serious violence against (or graphically depicts serious injury to) a person or animal (whether real or fictional); content that encourages, promotes, or provides instructions for stunts and challenges that are highly likely to result in serious injury; and content that encourages the self-administration of harmful substances.
[829] Non-designated content refers to content that is not primary priority content or priority content but presents a material risk of significant harm to an appreciable number of children in the UK. Ofcom's Children's Register of Risks identifies body stigma content and depression content as kinds of non-designated content.

Measure on recommender systems

11.8 Our June 2025 Consultation proposed that where service providers have 'relevant available information' which indicates that a piece of content is potentially illegal, providers should exclude it from content recommender feeds unless their content moderation function has reviewed it.[830] We did not propose that the content be taken down from the service entirely unless the provider has reviewed and assessed the content and determined that the content was illegal content and/or illegal content proxy.[831]

11.9 We proposed that providers of user-to-user services that have content recommender systems and are medium or high risk for at least one kind of the following kinds of illegal harms: hate, terrorism, encouraging or assisting suicide (or attempted suicide), or foreign interference offences (FIO) apply the measure. We refer to the types of harms providers are medium or high risk of as "relevant harms". We explained that recommender systems are a risk factor for exposing users to these harms and therefore proposed a new measure in our Illegal Content Codes.[832]

11.10 In our consultation, we considered that this measure would reduce the risk of users' exposure to, and engagement with, illegal content on recommender feeds. We also considered it would reduce the risk of illegal content going viral before moderation teams could review it.

11.11 Given the risks associated with recommender systems, we considered the direct and indirect costs of implementation were proportionate. We also considered any interference with user's rights to be proportionate to the benefits the measure would deliver in protecting users from the relevant harms.

Summary of stakeholder feedback

11.12 We received mixed feedback from stakeholders in response to the proposal in our June 2025 Consultation.

11.13 A range of stakeholders – representing civil society, industry, government, law enforcement, and academia – supported the proposed measure.[833]

---

[830] We used the term 'relevant available information' in a similar measure in our Protection of Children Code (Measure PCU E1: Content recommender systems: excluding potential primary priority content). We acknowledge that this term is similar to the term 'all relevant information that is reasonably available to a provider', which is a statutory term used in section 192 of the Act to refer to the information that providers should use where making a judgment about whether content is content of a particular kind (for example, making an illegal content judgement). However, we have decided to continue to use the term 'relevant available information' for consistency with measure PCU E1.

[831] Illegal content proxy refers to content that a provider determined to be in breach of its terms of service, where: a) the provider has reason to suspect that the content may be illegal content; and b) the provider is satisfied that its terms of service prohibit the type of illegal content which it had reason to suspect existed.

[832] Ofcom (2025), Illegal Harms Register.

[833] 5Rights Foundation response to the June 2025 Additional Safety Measures Consultation (June 2025 Consultation), pp.10-11; Antisemitism Policy Trust response to the June 2025 Consultation, p.3; Barnardo's response to the June 2025 Consultation, pp.13-14; British and Irish Law Education and Technology Association (BILETA) response to the June 2025 Consultation, p.19; CCDH response to the June 2025 Consultation, p.2; Check My Ads response to the June 2025 Consultation, pp.10-11; Children Commissioners Office (CCO) response to the June 2025 Consultation, pp.18-19; Commissioner Designate for Victims of Crime Northern Ireland response to the June 2025 Consultation, pp.6-7; Demos response to the June 2025 Consultation, pp.2-8; Department of Justice for Northern Ireland response to the June 2025 Consultation, pp.12-13; Evangelical

11.14 Other stakeholders requested clarity on the measure or raised concerns, which are outlined in the following paragraphs.

Freedom of expression and over-moderation concerns

11.15 A number of stakeholders – representing civil society, industry, law enforcement, and academia – raised various concerns about the measure, particularly in relation to its potential impact on users' rights to freedom of expression.[834] In particular, some industry stakeholders and individuals said that the concept of 'potentially illegal content' would introduce a new form of monitoring that would reduce users' freedom of expression and increase costs for providers.[835] Some stakeholders also expressed concern that terms like "potentially illegal" and "relevant available information" were overly broad and risked inviting overreach.[836]

11.16 Several stakeholders were concerned that the measure could result in over moderation, resulting in the excessive removal or suppression of lawful or potentially informative content from services.[837] Some respondents were especially concerned about relying on automated technology with limited human oversight and/or transparency, given the limitations of the technology.[838]

11.17 Additionally, various civil society, individual, and regulatory stakeholders raised concerns that the proposed measure did not mention notifying users who generate, upload, or share content that their content has been excluded from recommender systems, or providing them with the ability to appeal these decisions (the way users are able to appeal content moderation decisions that result in their content being taken down).[839]

Alliance UK response to the June 2025 Consultation, pp.9-10; Glitch response to the June 2025 Consultation, pp.14-18; [✂]; ICO response to the June 2025 Consultation, pp.30-31; [✂]; [✂]; National Police Chiefs' Council (NPCC) response to the June 2025 Consultation, pp.10-11; [✂]; National Society for the Prevention of Cruelty to Children (NSPCC) response to the June 2025 Consultation, pp.30-31; Online Safety Act (OSA) Network response to the June 2025 Consultation, pp.22-24; Refuge response to the June 2025 Consultation, pp.13-14; [✂]; [✂]; University College London (UCL) Gender and Tech Research Lab response to the June 2025 Consultation, p.3.

[834] Dr Are, C. response to the June 2025 Consultation, pp.15-17; [✂]; Francis, S. response to the June 2025 Consultation, p.13; Google LLC response to the June 2025 Consultation, pp.24-29; [✂]; [✂]; Klaushofer, A. response to the June 2025 Consultation, p.8; Integrity Institute response to the June 2025 Consultation, pp.7-8; Luethje, Y. response to the June 2025 Consultation, p.12; Dr Mack, S. response to the June 2025 Consultation, p.7; [✂]; [✂]; Name Withheld 4 response to the June 2025 Consultation, pp.7-8; Name Withheld 26 response to the June 2025 Consultation, pp.9-10; [✂]; Open Rights Group response to the June 2025 Consultation, pp.31-36; [✂]; Together Declaration Campaign response to the June 2025 Consultation, pp.3-6; [✂]; [✂]; [✂]; We Are Survivors response to the June 2025 Consultation, pp.7-8; [✂].

[835] Computer and Communications Industry Association (CCIA) response to the June 2025 Consultation, pp.12-13; Google LLC response to the June 2025 Consultation, pp.24-29; [✂]; [✂]; [✂]; [✂].

[836] BILETA response to the June 2025 Consultation, p.19; Francis, S. response to the June 2025 Consultation, p.13; Integrity Institute response to the June 2025 Consultation, pp.7-8; Luethje, Y. response to the June 2025 Consultation, p.12; Dr Mack, S. response to the June 2025 Consultation, p.7; Name Withheld 26 response to the June 2025 Consultation, pp.9-10; Together Declaration Campaign response to the June 2025 Consultation, pp.3-6; [✂].

[837] Google LLC response to the June 2025 Consultation, pp.24-29; [✂]; [✂]; [✂]; We Are Survivors response to the June 2025 Consultation, pp.7-8; [✂].

[838] [✂]; Glitch response to the June 2025 Consultation, pp.14-18; [✂].

[839] Dr Are, C. response to the June 2025 Consultation, pp.15-17; Demos response to the June 2025 Consultation, pp.2-3; [✂]; Name Withheld 4 response to the June 2025 Consultation, pp.7-8; Open Rights Group response to the June 2025 Consultation, pp.31-36; Online Safety Network response to the June 2025 Consultation, pp.22-24.

11.18 Some stakeholders suggested specific safeguards to protect users' right to freedom of expression, such as fast human review of flagged content or "downranking" flagged content instead of outright removing it.[840] One stakeholder said that a more proportionate approach would be to require services to enforce their demotion policies for content consistently and transparently rather than focusing on content removal.[841]

Scope and targeting of the measure

11.19 Google LLC suggested that the measure was applied too broadly. It said further that the measure expected services to apply the measure for all relevant illegal harms, even where a service is medium or high risk of only one kind of relevant illegal harm.[842]

Other stakeholder feedback

11.20 A number of stakeholders from industry and civil society provided general feedback to one of our questions relating to the manifestation of foreign interference campaigns. Some commented on and/or provided evidence related to existing practices of foreign interference (such as the use of bots to increase content virality), while others recommended caution in this area.[843] [844]

11.21 Other stakeholders requested changes to the measure, including an expansion of the harms in scope of the measure, expanding the measure to network recommender systems, making the measure more 'safety by design', and making allowances for livestreams and/or livestreaming-only services. We consider these responses in the relevant chapter in annex 1: Further stakeholder responses.

Our decision

11.22 In response to the feedback, we have decided to make changes to the measure we proposed in the June 2025 Consultation and outline our reasons for doing so in the following section and in the section titled 'Our reasoning'. The following changes are outlined in this section:

- Targeting the measures at children
- Removing the Foreign Interference Offence
- Adopting two measures in relation to types of harm

Targeting the measures at children

11.23 While we have provided evidence of significant, sustained, and systemic risk of harm to users from illegal content in recommender feeds, we are required to strike an appropriate balance between safety and freedom of expression. We have therefore decided to target the recommender system measures at children.[845]

---

[840] Glitch response to the June 2025 Consultation, pp.14-18; Name Withheld 26 response to the June 2025 Consultation, pp.9-10.
[841] Dr. Kira, B. response to the June 2025 Consultation, pp.6-10.
[842] Google LLC response to the June 2025 Consultation, pp.24-29.
[843] In the June 2025 Consultation, we asked respondents: 'Do you have evidence on what types of content are typically recommended to users as part of concerted foreign interference activity?'
[844] Antisemitism Policy Trust response to the June 2025 Consultation, p.3; Check My Ads response to the June 2025 Consultation, pp.10-11; [✂]; [✂]; Integrity Institute response to the June 2025 Consultation, pp.7-8; [✂]; Open Rights Group response to the June 2025 Consultation, p.35; [✂].
[845] We refer to 'measures' as we have decided to divide the measure on which we consulted into two measures. See explanation from paragraphs 11.38 onwards.

11.24 When content which has not yet been reviewed by a service's content moderation function is excluded from recommender feeds, there is a likelihood that a large volume of legal content will be excluded. This may significantly interfere with users' rights, including highly protected forms of expression such as political speech.

11.25 Given the relatively broad definition of 'content indicated to be potentially illegal content' that service providers should exclude from recommender feeds, and taking into account the stakeholder feedback received, we consider that the proposed measure would have posed disproportionate risks to adult users' freedom of expression.

11.26 However, we still consider it proportionate and appropriate to apply these measures to children's recommender feeds to protect them from content indicated to be potentially illegal content. Where we refer to 'the recommender feeds of children' or 'children's recommender feeds' in this chapter, we mean the content recommender feeds of users who have not been determined to be adults using highly effective age assurance.[846]

11.27 As part of this decision, we have decided these measures now apply:

a) to a provider in respect of each service likely to be accessed by children; and
b) where a child-accessible part of the service has a content recommender system.

11.28 The Online Safety Act 2023 (the Act) specifically makes provisions to ensure that children receive a higher standard of protection from online harms than adults.[847] We note in our Children's Register of Risks (Children's Register) that children are particularly vulnerable to harm as a result of exposure to harmful content, including the loss of life.[848] [849] [850] [851] We consider these concerns to be exacerbated where children encounter content encouraging or assisting suicide (or attempted suicide) or terrorism or illegal hate content. As we explain in more detail below, children's increased vulnerability to harm and the Act's provision that they should receive a higher standard of protection than adults, means that we consider the freedom of expression impacts of applying the measure to children to be proportionate.

11.29 The Government's announcement that it intends to restrict under 16s from using social media amongst other proposals could impact this measure. We nevertheless consider that this recommendation remains appropriate, because:

a) it is unlikely that all services in scope of these measures would be covered by such minimum age restrictions; and
b) under the Act, providers have duties to protect all children under 18 from both content harmful to children and illegal content, including 16- and 17-year-olds

---

[846] We refer to these users as 'relevant users' in the Illegal Content Code measure itself.
[847] Schedule 4 to the Act, paragraph 4(a)(vi).
[848] Ofcom (2025), Children's Register of Risks, p.299.
[849] The Coroner's report for 14-year-old Molly Russell concluded that it was likely that the material viewed by Molly – who was already experiencing a depressive illness, and was vulnerable due to her age – affected her mental health in a negative way and contributed to her death in a more than minimal way. Source: Courts and Tribunals Judiciary, 2022. Molly Russell: Prevention of future deaths report.
[850] Ofcom (2024), Experiences of children encountering online content promoting eating disorders, self-harm and suicide.
[851] Note: Research was with girls and women aged 15 to 25 22 countries, excluding the UK. Source: Plan International, 2020. Free to be online? Girls' and young women's experiences of online harassment; Note: Research was with 13 to 18-year-olds in the UK. Source: UK Safer Internet Centre (UKSIC), 2016. Creating a Better Internet for All: Young people's experiences of online empowerment + online hate.

11.30 Providers that expect to be in scope of the upcoming age restrictions should not delay the implementation of these proposed protections to their recommender systems, building on the existing Protection of Children measures PCU E1 and PCU E2.

11.31 We also consider that these measures build on our existing Protection of Children Code measures PCU E1 and PCU E2 and that there is likely to be substantial overlap between the relevant harms:

a) illegal content encouraging or assisting suicide (or attempted suicide) and primary priority content suicide and self-harm content; and,
b) terrorism or illegal hate content and priority content which encourages, promotes or provides instructions for an act of serious violence against a person, depicts real or realistic violence against a person, is abusive and targets race, or incites hatred based on race.

11.32 We consider that not extending these protections to relevant illegal harms would risk an inconsistent approach. Providers would otherwise be expected to apply a higher standard of protection for children in relation to content indicated to be potentially harmful to children (relevant primary priority content and priority content) than to content indicated to be potentially illegal content.

11.33 In alignment with our existing Protection of Children recommender systems measures, providers can choose to apply the measure(s) they are in scope of to the recommender feeds of all users if they do not wish to implement highly effective age assurance.

11.34 While we have decided to focus these measures on children's feeds at this time, recommender systems can also pose significant risks to adults and further work on the risks recommenders pose to adults will be required in future.

Removing the foreign interference offence

11.35 We have also decided to remove the foreign interference offence from the scope of the measures. We no longer consider it proportionate to include this offence.

11.36 While we maintain that recommender systems are a risk factor for the foreign interference offence, we now consider that the approach recommended by these measures could result in a significant volume of legal content being excluded from recommender feeds, before it has been determined to be illegal, if foreign interference remained in scope.

11.37 We consider that not making this change would lead to an increased risk of false positives occurring due to the challenges in attributing items of content to a foreign power, combined with the difficulty in distinguishing some foreign interference content from political speech. We consider that this could have a severe unintended impact which would represent a disproportionately negative impact on freedom of expression. Some respondents raised this as a material risk of our proposals, as identified in paragraph 11.108-11.110.

Adopting two measures in relation to types of harm

11.38 We have decided to divide the measure on which we consulted into two measures. Providers are therefore recommended to configure their systems and processes to only exclude the following groups of content indicated to be potentially illegal content where they assess themselves at medium or high risk of:

a) encouraging or assisting suicide (or attempted suicide) and serious self-harm;[852] and/or
b) terrorism and/or hate

11.39 For example, if an in-scope provider is only at medium or high risk for hate, it should implement the measure to exclude content indicated potentially to be illegal hate and/or terror, but need not implement the measure for encouraging or assisting suicide (or attempted suicide) and serious self-harm. Similarly, if it is only at medium or high risk for encouraging or assisting suicide (or attempted suicide) and serious self-harm, it should implement the measure to exclude content indicated potentially to be content encouraging or assisting suicide (or attempted suicide) and serious self-harm, but need not implement the measure to exclude content indicated potentially to be illegal hate and/or terror.

11.40 This is a change from our proposal in our June 2025 Consultation, where a service in scope of any of the relevant harms was recommended to implement the measure to exclude content indicated to be potentially illegal content for all of the harms in scope. We consider this is a more targeted approach, focusing on where there is a clear risk of harm. We further explain our reasoning for this decision in the section titled 'Who these measures apply to'.

How these measures work

11.41 The objective of these measures is to reduce children's exposure to relevant illegal harms on recommender feeds. To achieve this, providers should design and operate content recommender systems so that content indicated potentially to be illegal content is excluded from their feeds. This builds on existing measures in the Protection of Children Code measures PCU E1 and PCU E2.

Implementing the measures

11.42 To implement this measure, providers should take the following steps:

a) identify what relevant available information exists;
b) design the content recommender system to take appropriate account of that relevant available information; and
c) ensure the content recommender system operates so that content indicated potentially to be illegal content is excluded from children's recommender feeds.

11.43 This will protect children from being recommended and exposed to content that is indicated potentially to be illegal content, rather than waiting for it to be detected for content moderation and/or confirmed as illegal content or illegal content proxy through content moderation.

11.44 Excluded content which is then determined not to be illegal content, or an illegal content proxy, can be reinstated to recommender feeds (unless determined to be primary priority content, priority content, or non-designated content). Service providers should refer to our Protection of Children content moderation measures (PCU C1-2) for appropriate action.[853]

[852] As explained in chapter 1: Introduction, our duties and navigating the Statement, we consider encouraging or assisting suicide and serious self-harm as one kind of illegal harm.
[853] Protection of Children Code.

Relevant available information

11.45 By 'relevant available information' we mean any type of information that is reasonably available from the operation of the service and that a content recommender system can use to exclude content indicated to be potentially illegal content.

11.46 We expect service providers to have existing sources of relevant information and consider that a significant majority of in-scope providers will have already implemented the Protection of Children recommender systems measures (PCU E1 and PCU E2). We therefore expect these providers to have a clear understanding of which relevant available information they have and should use.

11.47 Providers should make use of the relevant information that is available to them, which may vary by service. Depending on how the recommender system is designed and operated, providers could use any of the following as relevant available information:[854]

a) content metadata;[855]
b) content labels or tags;[856]
c) data relating to the user uploading, generating or sharing content;[857]
d) user feedback[858] (including explicit negative feedback from children under measure PCU E3[859] or another negative feedback mechanism);
e) detection of content by automated tools;[860] and
f) user reports about illegality, including those from trusted flaggers.[861]

11.48 This is a non-exhaustive list of examples based on our understanding of relevant information typically available. Service providers may choose to introduce additional ways to gather relevant information that content may be illegal. We recommend that service providers ensure that any content recommender system on a child-accessible part of the service is designed and operated so that content that is indicated potentially to be illegal content is excluded from their feeds.

11.49 We consider this to be a precautionary approach. Providers should take this approach where their content moderation function has not yet determined the content to be illegal, or where the content may not be queued for content moderation. For example, they may train

---

[854] The definition of 'relevant available information' set out in our Illegal Content Codes includes 'indicators generated by technology used on the service'.
[855] Content metadata is the descriptive information about the content itself which can include various attributes that provide context, structure, and insights into the content, such as type (text, image, video), caption, hashtags, mentions, and engagement (likes, shares, and view counts).
[856] Content labels or tags are visual/textual information to categorise content, these may include moderation labels (such as misinformation, harmful, political), discoverability (such as sponsored, trending), and accessibility (such as trigger warnings).
[857] Data relating to the user uploading, generating or sharing content, such as whether other content uploaded, generated or shared by that user has been assessed to be illegal content.
[858] User feedback is the various types of data that helps the recommender systems learn about users' preferences, behaviour, and interactions with content.
[859] Protection of Children Code, Measure PCU E3.
[860] Where services detect content using proactive technology, this is an 'indicator generated by technology used on the service' and should therefore be treated as 'relevant available information' for the purpose of our recommender systems measures. We are recommending the use of perceptual hash matching to detect terrorism content (ICU C13) (see chapter 10: Hash matching).
[861] Trusted flaggers are individuals, non-governmental organisations (NGOs), government agencies, and other entities that have demonstrated accuracy and reliability in flagging content that violates a platform's terms of service. As a result, they often receive special flagging tools such as the ability to bulk flag content.

their systems using historical data containing relevant available information associated with identified illegal content, so that the system learns to not recommend content that has the relevant available information associated with potentially illegal content.

Our reasoning

Effectiveness and benefits

Risks these measures respond to

11.50 While we understand that users may benefit from recommender systems through the recommendation of content that they find interesting or want to see more of, we also recognise the risks associated with recommender systems. We understand that the content in children's recommender feeds is not necessarily moderated before it is recommended. Where a provider has detected content for moderation, there may be a time delay between detection and takedown (where appropriate).

11.51 This means that recommender systems can expose significant numbers of child users to potentially illegal hate, terrorism, and suicide and self-harm content before this content is detected, reviewed, and (where appropriate) taken down. This can cause harm both in the case of the initial exposure, as well as over time, through repeated exposure or through adversarial usage or 'gaming' of the recommender system design, as identified in paragraph 11.5.

11.52 As noted in paragraph 11.28, children are particularly vulnerable to harm resulting from exposure to illegal content. As noted by Jonathan Hall, K.C. the Independent Reviewer of Terrorism Legislation, "the internet is a major source of radicalisation and self-radicalisation of children, resulting in a well-documented increase in children being investigated and arrested for terrorism offences."[862]

11.53 The duties imposed on providers by the Act seek to secure that services are designed and operated in such a way that a higher standard of protection is provided for children than for adults.[863] Our Codes of Practice must be compatible with pursuit of the online safety objectives, including that the service provides a higher standard of protection for children than for adults and users should be protected from harm including in relation to algorithms and functionalities used by the service.[864]

Effectiveness of the measures

11.54 With these measures, we recommend providers take a 'precautionary approach', meaning that providers should exclude content from children's recommender feeds where it is indicated to be potentially illegal content, even if the content has not entered the moderation pipeline, or/if the provider's content moderation function has not yet determined content to be illegal content and/or illegal content proxy. Without these measures there is a risk that children will encounter illegal content on their recommender feeds.

11.55 We consider that taking this precautionary approach will reduce the extent to which children are exposed to illegal content, thereby delivering significant benefits, both by

[862] Independent Reviewer of Terrorism Legislation response to November 2023 Illegal Harms Consultation, p.1.
[863] Section 1 of the Act.
[864] Paragraph 4 of Schedule 4 to the Act.

reducing initial exposure and possible harm caused, and by reducing longer-term harm through continued or increased exposure.

11.56 These measures build on our Protection of Children recommender systems measures (see PCU E1 and PCU E2), ensuring that children receive additional protection from relevant illegal harms, as well as content harmful to children.

Relevant available information

11.57 We have outlined what we consider to be relevant available information in paragraphs 11.45-11.49. We have also included a non-exhaustive list of examples which we consider provides clarity to providers on what relevant available information we expect them to identify, and take appropriate account of in the operation and design of their recommender systems.

11.58 We expect much of the relevant available information used by providers in scope of these measures will already be used by providers in scope of the Protection of Children Code measures, PCU E1 and PCU E2.

11.59 Twitch considered that our definition of relevant available information was too broad. It said that temporarily removing content that is merely potentially illegal could result in a user with a single minor or accidental previous infraction having all of their content removed from recommender systems. This would have disproportionate impacts on users' rights.[865]

11.60 The British and Irish Law Education Technology Association (BILETA) also highlighted concerns regarding how the 'sources of relevant available information' indicate how likely content is likely to be illegal. It suggested this could create inconsistency in complying with the measure.[866]

11.61 We are not prescriptive about how service providers should use 'relevant available information' as recommender systems, moderation functions, and wider service design vary service to service. We consider services are therefore best placed to determine how to design and operate their content recommender systems to take account of all 'relevant available information' to meet the objective of the measures.

11.62 While 'relevant available information' is a broad concept, this is necessary to reflect the immense diversity of service designs and therefore the types and amounts of information available to service providers.

Potentially illegal content

11.63 In response to our June 2025 Consultation, two industry respondents raised concerns that the concept of 'potentially illegal content' is a new, undefined category of content which does not exist in the Act.[867]

11.64 Google LLC said further that Parliament decided that the decision not to extend the Act to content that was lawful but potentially harmful to adults was debated extensively in Parliament at the time of the Act's creation, and that there was legislative intent to exclude this content from the scope of the Act. For that reason, Google LLC considered that requiring service providers to identify 'potentially' illegal content is at odds with the intent of the Act

[865] Twitch response the June 2025 Consultation, pp. 19-22.
[866] BILETA response to the June 2025 Consultation, p.19.
[867] Google LLC response to the June 2025 Consultation, p.3; [✂].

(and out of step with other international regulatory regimes such as the EU's Digital Services Act (DSA)).[868]

11.65 Firstly, in regard to the concept of 'potentially illegal content', recommender systems commonly rely on a range of information and often recommend content before it has been moderated. We consider 'potentially illegal content' achieves the measures' aims of recommending that providers take a precautionary approach and exclude content from the feeds of users who have not been determined to be adults using highly effective age assurance, where there is relevant available information indicating that the content is potentially illegal content.

11.66 Secondly, while we agree that that the Act does not empower Ofcom to recommend measures in respect of 'legal but harmful' content for adults, we do not consider these measures target that category of harm. The measures are designed to protect children from content indicated to be potentially illegal content. This content relates to relevant harms where there is likely to be substantial overlap between these harms and content harmful to children, as explained in paragraph 11.31. We consider a precautionary approach of this nature is necessary to protect children effectively from illegal content appearing on their recommender feeds.

11.67 Some of the 'potentially illegal' content excluded will be content that the provider has reason to suspect may be illegal content. The service provider should review and assess such content and determine if the content is illegal content and/or illegal content proxy, and swiftly take it down.[869]

11.68 However, we accept that we are recommending that recommender feeds exclude content from children's recommender feeds on a precautionary basis, even if it is not a part of the content moderation queue. Our recommendation is that from the universe of millions or billions of items of content a provider could recommend, it avoids recommending to children content that in its experience has the relevant available information associated with illegal content. We recognise that this recommendation will affect legal content too and consider this to be an inevitable consequence of a precautionary approach. We discuss the implications that this measure has on rights in the section titled 'Rights impact' below. We consider the measures are appropriate and proportionate for protecting children from illegal content. While the content will be excluded from children's recommender feeds, we are not recommending the content is taken down from the service entirely.

11.69 We therefore do not consider the concept of 'potentially illegal content' unclear.

Freedom of expression

11.70 For a full consideration of these measures' impact on users' and providers' rights, please see the section titled 'Rights impact' below.

11.71 Some stakeholders suggested that our use of 'relevant information' and 'potentially illegal content' would ultimately lead to the exclusion of legal content from recommender feeds.

---

[868] Google LLC response to the June 2025 Consultation, pp.24-29.

[869] A stakeholder said that not all content excluded from recommender systems should be queued for content moderation. Source: [✂].

They said this would have a significant impact on both users' freedom of expression and service providers' moderation functions.[870]

11.72 This measure is designed to combat illegal content. However, we accept that as a by-product of the measure some lawful content may be excluded from users' recommender feeds, and that this impacts users' (and service providers') freedom of expression. In response to this potentially significant impact on users and providers, we have limited the focus of the measure to the recommender systems of users not determined to be adults. Children are more vulnerable to harm than most adults, and the Act necessitates a greater degree of protection for children. While we accept the measures will result in lawful content being excluded from children's recommender feeds, we consider this to be justified given children's heightened vulnerability to harm and the need to protect them from the viral spread of illegal content.

11.73 The measures do not go further than is necessary to secure that children are protected. They are focused on a limited range of harms, i.e., those where the Illegal Harms Register has identified content recommender systems as a risk factor.

11.74 The measures do not recommend the take down of such content from the service altogether (unless it is determined to be illegal content and/or illegal content proxy). Therefore, users will still be able to encounter the content by searching for it on the service or via a link, where these functionalities are available.

11.75 In many cases, content excluded under these measures may also be content indicated to potentially be content harmful to children and therefore remain unsuitable for recommendation to children. This content should therefore be excluded from children's recommender feeds under Protection of Children recommender measures PCU E1 and PCU E2.[871]

11.76 The recommendation to exclude the content no longer applies if, following content moderation, it is determined not to be illegal content (where content indicated to be potentially illegal content is initially excluded from feeds).

Notification of users and recourse to excluded content.

11.77 In our June 2025 Consultation proposals, we did not include a provision that providers give users the ability to have their content's exclusion from recommender systems reversed.

---

[870] Computer and Communications Industry Association (CCIA) response to the June 2025 Consultation, pp.12-13; Dr Kira, B. response to the June 2025 Consultation, pp.6-10; Google LLC response the June 2025 Consultation, pp. 24-29; [✂]; [✂]; [✂]; [✂].

[871] Relevant content harmful to children would be: i) Primary priority content (content which encourages, promotes, or provides instructions for suicide; content which encourages, promotes, or provides instructions for an act of deliberate self-injury); ii) Priority content: (content which is abusive and which targets any of the following characteristics: (a) race, (b) religion, (c) sex, (d) sexual orientation, (e) disability, or (f) gender reassignment; content which incites hatred against people (a) of a particular race, religion, sex or sexual orientation, (b) who have a disability, or (c) who have the characteristic of gender reassignment; content which encourages, promotes, or provides instructions for an act of serious violence against a person; or content which (a) depicts real or realistic, serious violence against a person, or (b) depicts the real or realistic serious injury of a person in graphic detail.)

Some respondents to our consultation suggested that, without appeals or notification of content's exclusion, the measure may result in users being 'shadowbanned'.[872] [873]

11.78 We have considered this feedback; however, we do not consider it technically practicable or proportionate to include in the measure that service providers should notify users every time their content is excluded from children's recommender feeds.

11.79 We understand that recommender systems exclude content in two ways:

a) Exclusions based on **explicit** recommendation decisions: these are based on predefined rules or conditions that directly determine which content should be excluded from the recommender feeds. Examples include excluding content that received negative feedback, content from blocked creators, content in restricted categories, or content that may violate business, trust and safety, or regulatory policies, e.g., by exceeding an internal content moderation threshold. These exclusions may be applied either by removing content from the candidate pool or by filtering it out during the final re-ranking stage. In either case, the exclusion reflects an explicit, traceable, and reversible business decision.
b) Exclusions based on **implicit** recommendation decisions: these do not rely on predefined rules or conditions that directly map signals to exclusion decisions. Instead, the recommender systems typically learn to identify and prioritise content for recommendation by learning user preferences alongside other provider priorities and objectives, such as user safety. In this context, content that is not selected for recommendation is implicitly left out (and by implication, excluded from the recommendation) because other content is selected. As a result, the non-recommendation decision is not deterministic and cannot always be traced back to a specific signal or rule, since it emerges from learned representation of user preferences, and other considerations, such as safety information.

11.80 While we acknowledge that some providers may only exclude content based on explicit exclusions, it is common practice for providers to exclude content based on both implicit and explicit exclusions. Where a provider uses explicit exclusions, whether exclusively or alongside implicit exclusions, it may be technically possible to notify users that their content has been excluded and offer them the opportunity to appeal the explicit exclusion.

11.81 However, where content is excluded as a result of an implicit decision, we do not consider that it is technically possible to notify users and provide an opportunity to appeal the decision. This is because providers may not know why a particular item of content has been excluded from a users' recommender feed as there is not an identifiable exclusion decision.

11.82 Further, we consider that notification of exclusions, whether implicit or explicit, could provide malicious users information necessary to design content which is more likely to bypass exclusion mechanisms and game recommender system designs. It may, therefore, undermine the effectiveness of the measure and have unintended consequences, leaving

[872] Shadowbanning is defined as "an act of a social media company limiting who can see someone's [content], usually without the person who has published [it] knowing." Source: Cambridge Dictionary, n.d. SHADOWBAN | English meaning - Cambridge Dictionary.
[873] Dr Are, C. response to the June 2025 Consultation, pp.15-17; Demos response to the June 2025 Consultation, pp.2-3; Glitch response to the June 2025 Consultation, pp.14-18; [✂]; Open Rights Group response to the June 2025 Consultation, pp.31-36; Online Safety Network response to the June 2025 Consultation, pp.22-24.

users less safe. This was noted by the Financial Conduct Authority (FCA) in its response to the consultation.[874]

11.83 We also consider that a notification system could result in large volumes of appeals. This may result in resources being reallocated from other systems and processes such as content moderation and other safety functions.

11.84 While we do not consider it feasible at scale to recommended notification or reversal of recommender exclusion, we recognise that many larger services offer tools to those users who are reliant on visibility, enabling them to monitor their content's virality.[875] These users are therefore likely to be able to see when their content has been affected by these measures. Such users would, under our measures on complaints by UK users and affected persons, be able to complain if they consider that a provider is not complying with its safety duties or its duties in relation to freedom of expression or privacy.[876]

Conclusion

11.85 Overall, we consider that these measures will provide significant benefits for children by reducing the risk of them accidentally or unexpectedly encountering relevant illegal content, by ensuring less of it will be present in children's recommender feeds.

11.86 These measures also aim to mitigate the risk that children are recommended more relevant illegal content after seeking it out and therefore will support the protection of children from more long-term harms such as those described in paragraph 11.4.

Impacts and costs

11.87 This section considers the direct and indirect costs of the measures and has been updated to reflect the implications of our changes since the June 2025 Consultation.

Direct costs

11.88 In order to exclude content indicated potentially to be illegal content from the recommender feeds of users, providers may incur costs related to:

a) identifying what relevant available information exists;
b) designing the recommender system to take appropriate account of that relevant available information; and
c) ensuring the content recommender system operates so that content indicated to be potentially illegal content is excluded from children's recommender feeds.

11.89 In our consultation, we noted that the costs of implementing the measure would depend on factors including the design and complexity of a service's recommender systems and the number of recommender systems operating on a service. We considered that costs would be higher where providers operate multiple recommender systems, and where systems are more complex; for example, systems serving larger numbers of users and operating across multiple languages. Providers which do not already have mechanisms to limit the prominence of relevant content types would face higher costs. Many providers have also already designed their recommender system(s) to avoid recommending certain types of

[874] Financial Conduct Authority (FCA) response to the June 2025 Consultation, p.4.
[875] We note examples of this practice identified by Instagram, n.d. About Instagram insights | Instagram Help Center; TikTok, n.d. https://www.tiktok.com/tiktokstudio/analytics; and YouTube, n.d. How engagement metrics are counted - YouTube Help.
[876] Illegal Content Codes, Measures ICU D1, D2 and D12.

content; for these providers, the implementation of these measures is likely to be more straightforward.

11.90 Several stakeholders supported our assessment of impacts and/or costs,[877] while others disagreed with our assessment of direct costs.[878] Some stakeholders said further that the measure would increase the demands on services' content moderation teams, potentially limiting their ability to remove harmful content.[879]

11.91 However, we have not seen evidence that would justify revising our cost estimates. We have updated our costs assessment to reflect the change we have made to these measures to use highly effective age assurance to target it at children and to use the latest Office for National Statistics (ONS) salary data.[880] Beyond these updates, our assessment remains as set out in our June 2025 Consultation.

11.92 We expect the vast majority of services in scope of these measures will incur only small incremental costs. This is because we consider that most services will already be in scope of our equivalent measures in the Protection of Children Code (PCU E1 and PCU E2), and these services will already have systems in place to exclude content harmful to children from recommender systems and will only need to adapt those systems to also address the relevant illegal harms.

11.93 For relevant service providers not in scope of our equivalent Protection of Children Code measures we estimate they could incur one-off costs between £17,000 and £102,000 to implement both measures taken together, and ongoing maintenance costs of between £4,000 and £26,000.[881] Services only in scope of one of the measures may incur lower costs. If services choose to use highly effective age assurance to target the measures at children, they may incur additional costs set out in paragraphs 11.98-11.102.

11.94 We acknowledge the stakeholder responses that raised concerns about the potential burden the measures could place on content moderation teams. However, these measures do not change the legal threshold above which the Codes of Practice stipulate that service providers should review content, and therefore do not in themselves increase content moderation costs. If a service chooses to moderate content to be able to recommend this content again, this would be a discretionary business decision. We recognise services may choose to take such an approach to mitigate any indirect impact on user engagement arising from excluding content from recommender systems. Doing so could increase the volume of content subject to moderation and consequently increase moderation costs. However, because these costs

---

[877] BILETA response to the June 2025 Consultation, pp. 20-21; Department of Justice for Northern Ireland response to the June 2025 Consultation, pp.12-13; Evangelical Alliance UK response to the June 2025 Consultation, pp.9-10; [✂]; Online Safety Network response to the June 2025 Consultation, pp.22-24; Which? response to the June 2025 Consultation, pp.11-13.

[878] [✂]; [✂]; Together Declaration Campaign response to the June 2025 Consultation, pp.3-6; UK Finance response to the June 2025 Consultation, pp.12-13.

[879] Google LLC response to the June 2025 Consultation, pp.24-29; Integrity Institute response to the June 2025 Consultation, pp.7-8; [✂]; [✂]; Open Rights Group response to the June 2025 Consultation, pp.31-36; [✂]; [✂]; [✂].

[880] See annex 5: Further detail on economic assumptions and analysis.

[881] This estimate is based on our assumption of 6 to 18 weeks of software engineering time, alongside an equivalent amount of time from staff in non-technical professional roles, such as lawyers, analysts, and product managers. In line with our standard cost assumptions, ongoing maintenance costs are estimated at 25% of the initial setup costs.

would result from optional actions taken at a service's discretion, we have not included them in our impact assessment for these measures.

11.95 Splitting the measure into two means that services will only need to apply these measures to the kind(s) of illegal content for which they assess themselves as medium or high risk. As a result of this more targeted approach, we consider the cost of implementing these measures will be lower, as a service will need to design a system to address fewer harms, unless they are medium or high risk for a relevant harm in both groups. If a service does choose to moderate some or all the content excluded from recommended feeds, limiting the measures to harms for which they are medium or high risk will reduce any associated content moderation costs.

Direct costs related to age assurance

11.96 We have changed the measures so that services have the flexibility to either apply the measures only to users not determined to be adults by use of highly effective age assurance or apply the measures to all users.

11.97 For services that choose to apply the measures to all users, we assess there will be no change in the cost assessment, aside from the updates made to reflect the latest ONS salary figures.

11.98 As we expect the vast majority of service providers to already be in scope of the Protection of Children recommender systems measures, they will have either implemented highly effective age assurance already or chosen to apply the relevant measures to all users. Where providers apply the same highly effective age assurance to both the Protection of Children's measures and the measures in this statement, we consider they will incur minimal incremental costs.

11.99 In the small number of cases where service providers are in scope of these measures and are not already applying highly effective age assurance for the purposes of compliance with another measure, they may incur more significant costs. Services can choose either to develop highly effective age assurance internally or to use a third-party provider. Where services chose to use an external provider, we estimate costs of between £0.05 to £0.30 per age check. We set out indicative age assurance costs for providers of services of different sizes in annex 5: Further detail on economic assumptions and analysis. For example, for a service with 100,000 users and 10,000 new users each year, there could be a cost of £5,000 to £30,000 to conduct age assurance checks for all existing users, and a cost of £1,000 to £3,000 annually for new users. For a service with 7,000,000 users and 70,000 new users each year, the one-off costs could be £350,000 to £2,100,000, and annual costs of £4,000 to £21,000.

11.100 Service providers have flexibility in their use of highly effective age assurance, so they do not need to check the age of every user.[882] For example, the measures may initially be applied to all users' recommender feeds. Users who subsequently verify that they are adults using highly effective age assurance can then access a feed from which content indicated to be potentially illegal in relation to the relevant harms is not excluded.

11.101 There may also be costs associated with integrating the age information from the age assurance process into the recommender system, so that children are shown a

[882] Costs in paragraphs 11.99-11.101 are likely to represent an upper bound. The age assurance costs could be significantly lower if age checks are only performed on a subset of users.

recommender feed with the relevant potentially illegal content excluded. We estimate that the one-off cost of this could be between £11,000 to £45,000, with annual maintenance costs approximately between £3,000 to £11,000. Again, for services already in scope of the equivalent measures in the Protection of Children Code, we consider providers will only incur a small incremental cost as they would likely already have linked this information.

11.102 There will also be some limited costs to update relevant parts of the user interface or settings related to the interaction between age assurance and the recommender system, for example, messages to inform users that they need to complete an age check to have access to a recommender feed that does not have these measures applied.

Indirect costs

11.103 As set out in our June 2025 Consultation, these measures may impose indirect costs on service providers and on individuals who earn income through the service if it reduces engagement and, in turn, revenue. This includes both individuals who monetise content directly through the service and those who use the service to promote offline businesses.

11.104 Providers have incentives to retain as much content on their services as possible to support user engagement. They may do this by improving the accuracy of the relevant available information and systems they use, reducing the risk that legal content is excluded. We also consider that targeting the measures to children may reduce the indirect impact of these measures, as the change to recommender feeds can be applied to a subset of users.

11.105 We recognise that some content which is legal and may have been engaging to users may also not be recommended to users because of these measures, but it can be reinstated if it is determined not to be illegal content and/or illegal content proxy via the content moderation process. For service providers, we expect this impact to be limited because they typically have a wide variety of other content they can recommend instead.

11.106 There may also be countervailing benefits of these measures to service providers where the measures decrease children's exposure to illegal content and users feel safer. This may mean these users spend more time on such services, or advertisers may feel more comfortable to advertise next to recommended content.

11.107 We did not receive stakeholder feedback relating to our assessment on the indirect costs of the measure at consultation and so our assessment remains unchanged.

Rights impact

Freedom of expression

Feedback from stakeholders

11.108 Many of the concerns we have addressed in paragraphs 11.23-11.86 have human rights implications. Concerns about whether the measures are clear enough, in paragraphs 11.57-11.69, relate to whether the measures are sufficiently foreseeable to be prescribed by law or in accordance with the law. Concerns about impacts on lawful speech, paragraphs 11.70-11.76, relate to whether the measures are proportionate.

11.109 A number of consultation respondents expressed concern that the measure on which we consulted could have the effect of chilling free speech.[883] Some stakeholders were particularly concerned that some of the most protected forms of speech could be

[883] [✂]; Google LLC response to the June 2025 Consultation, p.29; [✂]; Name Withheld 4 response to the June 2025 Consultation, pp.7-8; [✂]; [✂]; [✂].

disproportionately affected, for example journalistic content, political expression.[884] Some respondents also expressed a concern that the measure could be utilised to suppress legitimate speech from users.[885]

11.110 Some stakeholders also expressed concern about the risk that automated moderation would disproportionately affect marginalised communities.[886]

Is there an interference with freedom of expression?

11.111 We have amended the measure in the light of these comments, but still recognise that exclusion of users' content from the recommender feeds of users not determined to be adults is an interference with their rights to freedom of expression. It affects both their right to impart and to receive information. It also impacts on the rights of providers to promote the content they wish. Content will, of course, only be excluded from recommender feeds and not the whole service, meaning that the content is still discoverable and can still be forwarded or cross posted to other services where that functionality is available. However, recommender feeds are an important way for users to discover content.

11.112 The extent of the impact is impossible to assess as it is not possible to know what content would have been surfaced but for the measures. For the purpose of our rights assessment, we have assumed a very large amount of lawful content could be impacted because the approach we are recommending is precautionary. We acknowledge both that the content affected may include highly protected forms of expression such as political speech, and that it may affect content which engages other rights (for example, there may be a commercial impact for some users).

Prescribed by law

11.113 While we have provided examples of the types of information which could be relevant available information for the purpose of the measure, we acknowledge that the measures do not prescribe exactly what relevant available information a provider should rely on or what level of assurance it should apply. However, the measures are clear both about the objective to be achieved – children should not be exposed to illegal content on recommender feeds, and about the method – providers should take content out of children's feeds when they have relevant available information that that it may not be safe. Users may not know precisely what relevant available information a given service is using, but can foresee that their content may be taken out of children's feeds if it has the same relevant available information as illegal content. The harms concerned are public, and users are likely to be aware if they have previously been the subject of upheld complaints because their content will have been taken down.

11.114 In light of stakeholder feedback, we considered whether we could bring greater specificity and prescription to the measures. However, we decided that it was neither practicable or appropriate to do so, for the reasons set out at paragraphs 11.78-11.86.

---

[884] Google LLC response to the June 2025 Consultation, p.29; Open Rights Group response to the June 2025 Consultation, pp.31-36.
[885] [✂]; Francis, S. response to the June 2025 Consultation, p.13; [✂]; Klaushofer, A. response to the June 2025 Consultation, p.8; Open Rights Group response to the June 2025 Consultation, pp.31-36; Together Declaration Campaign response to the June 2025 Consultation, p.3.
[886] Dr Are, C. response to the June 2025 Consultation, pp.15-17; [✂].

Necessary for a legitimate purpose

11.115 As set out in paragraphs 11.3-11.5 the risks to children and others from children's exposure to illegal content in their recommender feeds are significant. Exposure to terror and hate content gives rise to risks of crime and disorder, including risks to national security and to public safety. Exposure to suicide and self-harm content gives rise to risks to public safety and health.

11.116 Users' rights to freedom of expression must be balanced against the rights of users not to be drawn into terror or hate, not to be forced to experience hate, and not to be encouraged or assisted to hurt themselves.

11.117 We do not consider that users not determined to be adults can be protected effectively by waiting until a provider is sure that content is illegal content, before excluding it from feeds. This is because of the speed with which harms can spread when a recommender system contributes to the rapid increase in the number of UK users encountering the content. Prior to the advent of the internet, and in particular the invention of recommender systems, it was not possible for harms to achieve such scale and to reach children directly in their homes/everywhere, such that there was no need to intervene in advance of having decided that content was illegal. The same is not true now: the harm cannot be addressed without rapid precautionary action against content. Nor do we consider that merely downranking the content would necessarily address the risks, since content may only be downranked a little.

11.118 We do not consider that a higher level of assurance for taking content out of feeds would be likely to be effective at protecting children. Content on recommender feeds can go viral so rapidly that there is a need for intervention before signals have had time to accumulate. Our measures ask service providers to design and operate their recommender systems with a view to preventing illegal content from being promoted to children. We consider that they have a strong commercial incentive to implement this in the way that has the least possible impact on the volume, range and diversity of lawful content they can promote, based on the information they have.

11.119 As such, we consider the measures to be the minimum necessary to achieve the legitimate purposes.

Proportionate

11.120 We recognise that the measures may have significant impacts on lawful expression including political expression, religious expression, creative expression and expressions of significant commercial value. Recognising the impacts on freedom of expression, we are only recommending these measures for the feeds of users who have not been determined to be adults. The duties imposed on providers by the Act seek to secure (among other things) that the services they provide are safe by design and designed and operated in such a way that a higher standard of protection is provided for children than for adults.[887] While children have a right to receive information, their particular vulnerability means that a greater interference with the right may be proportionate, than it would be for adults. In cases where it is not certain whether a user is an adult or a child, the vulnerability of children and the severity of the harms means we consider it is better to risk treating adults as children, than children as adults.

[887] Section 1(3) of the Act.

11.121 We note that all users will still have other ways (both on and offline) to receive and impart the affected content and to discover the content.

11.122 Providers are not required, and have no incentive under these measures, to target expressions of a particular political opinion or by a particular group. On the contrary, failing to detect and manage bias in their systems would be likely to have significant reputational consequences and could expose them to risks under equality or data protection legislation. As such, we consider this risk to be minimised.

11.123 We also note that providers are likely to have an incentive to maximise the volume, range and diversity of lawful content they can promote, based on the information available to them. The measures should therefore reinforce incentives they have to improve the accuracy of the relevant available information and systems they use to identify content that that can be safely recommended.

11.124 In this statement, we are recommending the use of hash matching to detect terrorism, which is likely to generate relevant available information for a recommender system. Our recommendation includes safeguards in relation to quality assurance, content moderation and associated appeal rights, which are set out in chapter 10: Hash matching.

11.125 Finally, users can still complain to service providers that they are not complying with the safety duty by applying their terms and conditions consistently, or that they are not complying with their duties in relation to freedom of expression.

11.126 Overall, we consider that in relation to freedom of expression, the measures are the least intrusive means possible of protecting children from very significant harm, and are not disproportionate.

Privacy

Feedback from stakeholders

11.127 An individual stakeholder [✂].[888] By contrast, the ICO positively assessed the measure's alignment with minimal personal data protection.[889]

Is there an interference with the right to privacy?

11.128 We recognise that in order to determine how the relevant information available to a service provider correlate with illegal content, the provider is likely to need to consider a wide range of information which engages a user's right to private and family life. The relevant available information which prove valuable in protecting users who may be children from illegal content would be processed on an ongoing basis and may require ongoing surveillance of its users by the service provider. There are associated risks that the information collected may be disclosed to third parties, for example if the service provider is hacked. There is an interference with the right to private and family life.

11.129 Where a service provider chooses to use highly effective age assurance, this raises privacy issues. These are considered in chapter 5: Highly effective age assurance in the Illegal Content Codes.

[888] [✂].
[889] ICO response to the June 2025 Consultation, pp.30-31.

In accordance with the law

11.130 Service providers implementing these measures will be subject to both data protection and equality legislation and will be securing their compliance with the Online Safety Act. We therefore consider that the measures are in accordance with the law.

Necessary for a legitimate purpose

11.131 For the reasons set out in paragraphs 11.115-11.119 in relation to freedom of expression, we consider the interference with the right to private and family life to be necessary to achieve a legitimate purpose.

Proportionate

11.132 The measures do not recommend the collection or use of user data that is not already available to the provider, and we note that in order to predict the preferences of users any user choosing a service which operates a recommender system is already consenting to a significant degree of collection, and use, of the data so obtained. Providers subject to the measures are also subject to data protection legislation and equality legislation which we consider will minimise the degree of interference with privacy.

11.133 The privacy and data protection considerations relating to the use of highly effective age assurance are addressed in chapter 5: Highly effective age assurance in the Illegal Content Codes, which sets out when and how highly effective age assurance may be used for the purposes of applying additional safety measures, including those relating to recommender systems. Where a service provider is using highly effective age assurance, our Codes of Practice set criteria of technical accuracy, robustness, reliability and fairness which safeguard users' privacy.

11.134 Overall, we consider that the measures are the least privacy-intrusive means possible of protecting children from very significant harm, and that in a context where users have chosen an inherently unsafe service on which to discover content, it is not disproportionate.

Who these measures apply to

11.135 In our June 2025 Consultation, we proposed that the measure should apply to providers of user-to-user services that have a content recommender system and are medium or high risk for at least one of the following kinds of priority illegal content: hate, terrorism, encouraging or assisting suicide (or attempted suicide) and/or foreign interference offence content.[890]

11.136 We considered that there is sufficient evidence to identify recommender systems as a significant risk factor for those harms, as set out in our Risk Assessment Guidance.[891] We assessed that the measure would create significant benefits by helping to prevent users from encountering relevant illegal content of those kinds of harm on their recommender feeds.

11.137 We have made three changes to the application of these measures because of stakeholder feedback, outlined in turn in the following paragraphs.

---

[890] As explained in chapter 1: Introduction, our duties and navigating the Statement, once the Codes of Practice come into force encouraging or assisting suicide (or attempted suicide) and serious self-harm are to be considered one single kind of illegal harm.

[891] Ofcom (2024), Risk Assessment Guidance and Risk Profiles, p. 40-41.

Change to recommending the measures only to services likely to be accessed by children

11.138 As already discussed, having considered stakeholder feedback we have decided to target the measures at children. We have changed these measures so that they are now only recommended for services that are likely to be accessed by children according to a service's children's access assessment.[892] This aligns the application of these measures to the equivalent recommender systems measures in the Protection of Children Code. We consider that this is a proportionate way to target the measure at services with children on them while not creating additional work for services by mirroring the application in the Protection of Children Code measures.

Splitting into two measures for specific groups of harm

11.139 Google LLC stated that the measure was poorly targeted because it expected services to exclude all relevant illegal content, even where a service has assessed as medium or high risk of only one kind of illegal content.[893] In response to this feedback, we have decided to split the measure consulted on into two separate measures.

11.140 We recommend that services implement the measure(s) only for the kind(s) of harms for which they are at medium or high risk. This means we are recommending that services assess relevant available information and take action to exclude content only for those groups of harms assessed as medium or high risk. For example, a service assessed as high risk only for encouraging or assisting suicide (or attempted suicide) and serious self-harm content would not be expected to implement the measure in relation to content indicated to be potentially illegal hate or terrorism if it has assessed as low risk for those harms.

11.141 We consider it appropriate to group hate and terrorism harms into one measure as these harms are closely linked and often appear together.[894]

11.142 This approach aligns with our approach in the Protection of Children Code, in that the harms in scope of ICU E3 correspond with the relevant forms of primary priority content addressed under measure PCU E1, while those in scope of ICU E2 correspond with those addressed under PCU E2. We explain this further in paragraphs 11.31 and 11.75 above.

Change in the harms in scope of the measures

11.143 We are no longer recommending these measures for services that are medium or high risk of the foreign interference offence. Our rationale for this is set out in paragraph 11.35-11.37. This change means that services with a recommender system that are at medium or high risk of foreign interference offence content, but not at medium or high risk of any other relevant harms, are no longer in scope of these measures.

Proportionality assessment

11.144 We have decided that it is proportionate to apply these measures to service providers of any size where they are medium or high risk of the relevant harms. As set out in the Illegal

---

[892] Ofcom (2025), Children's access assessments. A children's access assessment is a process for establishing whether a service is likely to be accessed by children.
[893] Google LLC response to the June 2025 Consultation, pp.24-29.
[894] Note: a commissioned report to build Ofcom's understanding of user experiences of online services identified there is "significant overlap between a broad spectrum of harmful conspiracists and overtly white nationalist communities online. Disinformation, conspiracy theories, targeted hate, harassment, and extremism often interlink online in ways that are extremely difficult to separate and isolate". Source: Institute for Strategic Dialogue (ISD), 2023. Tangled Web - The interconnected online landscape of hate speech, extremism, terrorism and harmful conspiracy movements in the UK.

Harms Risk Profiles, services with content recommender systems are likely to have increased risk of hate, terrorism, suicide and self-harm, unless there are significant countervailing factors reducing that risk.[895] This could include comprehensive systems and processes that limit the exposure of users to this content.

11.145 We recognise that the estimated costs of these measures could be significant, including for some smaller service providers. This may result in a potential loss of user choice if a provider chooses to remove their recommender systems (and instead just provide a chronological feed) or to withdraw from the UK market entirely. This would mean fewer services for consumers to choose from. However, the measures only apply to services at medium or high risk of the relevant harms and do not apply to small, low risk services. We consider that excluding smaller services from scope altogether would risk leaving a gap in protection for children where those services still present relevant risks.

11.146 We consider that the costs are proportionate given the material benefits associated with these measures. This is supported by the fact that we generally expect costs to scale with the size and complexity of services; larger services typically recommend far greater volumes of content and, as a result, are likely to have more content indicated to be potentially illegal content to exclude.[896] Providers of smaller services are therefore likely to incur lower costs than larger services. We also consider these measures allow service providers appropriate flexibility in how they are implemented.

11.147 We are therefore recommending these measures for providers of user-to-user services that are likely to be accessed by children, have a content recommender system on a child-accessible part of the service, and are medium or high risk for at least one of the following kinds of illegal harms:

a) hate, terrorism (ICU E2); and/or
b) encouraging or assisting suicide (or attempted suicide) and serious self-harm (ICU E3).

Conclusion

11.148 We have decided to focus the measures on children rather than adults. This is because, for the reasons set out in the chapter, children require greater protection than adults.

11.149 Overall, we expect the vast majority of services in scope to incur only small incremental costs, as most are likely to already have systems in place to implement the equivalent Protection of Children measures. Costs may be more significant for the smaller number of providers that do not. The measures could also interfere with children's right to freely receive and impart certain information.

11.150 However, on balance we consider these costs and impacts on rights to be proportionate, given the magnitude of the risks to children posed by algorithmic dissemination and amplification of illegal content and the important role this measure will play in addressing these risks. As such, we have concluded that the measures would be proportionate, even in light of the costs to providers and the potential impact on children's rights.

11.151 We have therefore decided to recommend the measures in our Illegal Content Codes of Practice, where they are referred to as ICU E2 and ICU E3.

---

[895] Ofcom (2024), Risk Assessment Guidance and Risk Profiles.
[896] The impact of the size and complexity of the service on the cost is discussed in more detail in 'Impacts and Costs', paragraph 11.87.

12. Amendments to the Illegal Content Judgements Guidance for child sexual abuse material

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

What is this chapter about?

Perpetrators use a range of online services, such as messages, group chats or forums, to facilitate the commission of offences involving child sexual abuse material (CSAM) and to share CSAM. In some cases, perpetrators of CSAM offences use the name, icon or bio / description on these services to indicate to other perpetrators that CSAM content can be found on part of the service (e.g. group chat or forum). This type of user data may be useful for service providers to identify perpetrators of CSAM on their services, particularly if they are unable to review the content of the users' messages, group chat or forum. This chapter sets out guidance to support service providers in assessing these types of information in order to determine whether it is reasonable to infer that a messaging thread, group chat or forum exists for the purposes of committing CSAM offences.

What decisions have we made?

We have amended the Illegal Content Judgements Guidance (ICJG) to support service providers in making judgements about illegal content when they are unable to review messages, group chats, or forums.

To do this, we have set out three scenarios that illustrate how a provider can infer that content is illegal and explain our reasoning in this chapter.

Why have we made these decisions?

These scenarios will support service providers to identify and reduce the spread of CSAM on their services. As a result, this guidance directly contributes to the protection of survivors and victims of child sexual abuse.

Introduction

12.1 The Online Safety Act 2023 (the Act) requires providers of regulated user-to-user services to take certain steps to reduce the risk of harm to users from illegal content including child sexual abuse material (CSAM).

12.2 Section 192 of the Act sets out the approach that providers of in scope services must take when they are required to make a judgement about whether content is illegal content. In making such judgements, the approach to be followed is whether they have "reasonable grounds to infer" it is illegal content (that is, that it "amounts to a relevant offence").

Providers must use all relevant information reasonably available to them ("reasonably available information") to make this judgement.[897]

12.3 Ofcom has produced guidance for providers on how they can make illegal content judgements for the purposes of the takedown duty, the risk assessment duty, and the safety duties more generally (the Illegal Content Judgements Guidance (ICJG)).[898] The updated version of the ICJG was published alongside this statement.

Child sexual abuse material in messaging, group chats and forums

12.4 We recognise the significant harm caused by CSAM. Evidence suggests that perpetrators use messaging, group chats and forums to facilitate the commission of CSAM offences including by exchanging CSAM, links, advice and tradecraft tips.[899] The National Society for the Prevention of Cruelty to Children (NSPCC) highlighted the spread of CSAM within online communities and further noted that such behaviour is "normalised or even encouraged" due to their membership of like-minded individuals (i.e., those with a sexual interest in children).[900] Similarly, Protect Children also found open (and dark) web forums, chatrooms and messaging apps being used to either search for or share CSAM, which may lead, intentionally or unintentionally, to users of those services being exposed to CSAM.[901] We consider this evidence suggests that CSAM may be accessible to users of these services, regardless of whether they are actively searching for such content.

12.5 There is also evidence to suggest that service providers are not reporting cases of CSAM on their services to the appropriate authorities because it is not being detected. The National Centre for Missing & Exploited Children (NCMEC) experienced a decrease in the number of child sexual exploitation reports from 36.2 million reports in 2023 to 20.5 million in 2024.[902] It attributed this decrease to detection challenges for service providers, citing specifically that end-to-end encryption is increasingly used which significantly limits the content that service providers are able to detect and report.[903]

12.6 We consider the detection and reporting of CSAM, against the backdrop of the increasing use of end-to-end encryption, to be a significant challenge for service providers. However, even where messages are end-to-end encrypted, information may be available to the provider about the purpose of, or type of, content shared within the chat or forum. For example, perpetrators may use their online name, icon or bio/description on a providers' service to indicate to other perpetrators that CSAM content can be found within message threads, group chats or forums. For some service providers, it may be technically feasible to review that information but not the content of the messages, group chat or forum.[904]

---

[897] Ofcom (2024), Illegal Content Judgements Guidance, 'Illegal content judgements: Reasonable grounds to infer' and 'Reasonably available information', p.11-16.
[898] Section 193 of the Online Safety Act 2023 (the Act) requires Ofcom to produce this guidance.
[899] Ofcom (2024), Illegal Harms Register of Risks (Illegal Harms Register), p.95-96.
[900] National Society for the Prevention of Cruelty to Children (NSPCC), 2019. Online abuse: learning from case reviews.
[901] Protect Finland (2026), Tell Me More About Tech: CSAM Perpetrator Research Report, Findings from a survey of child sexual abuse material (CSAM) perpetrators on digital platform use and design, pp.15, 19-20.
[902] National Center for Missing & Exploited Children (NCMEC) (2025), 'NCMEC Releases New Data: 2024 in numbers'.
[903] NCMEC (2025), 'NCMEC Releases New Data: 2024 in numbers'.
[904] Ofcom (2025), June 2025 Additional Safety Measures Consultation, p.123-124.

12.7 In this chapter, we set out the final recommended amendments to our ICJG for CSAM to support service providers in making illegal content judgements in cases where the content is not accessible to the service provider. This includes detailing stakeholder feedback and our reasoning.

Our proposal

12.8 In the June 2025 Additional Safety Measures Consultation (June 2025 Consultation), we proposed changes to the ICJG for CSAM. The proposed amendments set out guidance for providers on how they could make illegal content judgements based on publicly available information.[905]

12.9 We identified two scenarios which we provisionally considered would give service providers reasonable grounds to infer that the content within the messaging, group chats or forums in question is CSAM:

i) Scenario one: The image or icon for the message, group chat or forum is CSAM; the image or icon for the messages, group chat or forum is a still from a known CSAM video; or the bio or heading for the message, group chat or forum explicitly indicates that the contents are CSAM.
ii) Scenario two: The bio or heading relating to a message, group chat or forum indicates that the content of the chat is CSAM through coded terms strongly associated with CSAM. However, providers should consider this content in combination with other available information, such as user complaints or police reports.

Summary of stakeholder feedback

12.10 Several stakeholders supported our proposed amendments to the guidance.[906] This included civil society stakeholders, a law enforcement agency and a government stakeholder.

12.11 Some stakeholders also supported our assessment of harm caused by CSAM and referred to further evidence that corroborated our view of perpetrator behaviour, as set out in the June 2025 Consultation.[907] Specifically, the Internet Watch Foundation (IWF) provided evidence which highlighted a public group on an end-to-end encrypted messaging app that used

[905] In our June 2025 Consultation, we explained that a 'signal' could be an image, icon, bio or heading which can sometimes be an indicator to perpetrators that the contents of messages, group chats or forums contain CSAM or are being used to facilitate the commission of CSAM offences. See Ofcom (2025), Annex 11: Proposed amendments to the Illegal Content Judgements Guidance, paragraphs 4.64 to 4.70, for more detail.
[906] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.16; Barnardo's response to the June 2025 Consultation, p.10; Canadian Centre for Child Protection (C3P) response to the June 2025 Consultation, p.11; Children's Commissioner's Office (CCO) response to the June 2025 Consultation, p.15; Digital Resilience in Education (Welsh Government) response to the June 2025 Consultation, pp.9-10; Evangelical Alliance response to the June 2025 Consultation, p.6; Lucy Faithfull Foundation response to the June 2025 Consultation, p.5; [✂]; [✂]; NCMEC response to the June 2025 Consultation, p.4; Online Safety Network response to the June 2025 Consultation, p.16; Social Media Animal Cruelty Coalition (SMACC) response to the June 2025 Consultation, p.10; We Are Survivors response to the June 2025 Consultation, p.3.
[907] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.16; C3P response to the June 2025 Consultation, p.11; Evangelical Alliance response to the June 2025 Consultation, p.6; Internet Watch Foundation (IWF) response to the June 2025 Additional Safety Measures Consultation, p.14; 26-27; National Police Chiefs' Council (NPCC) response to the June 2025 Consultation, p.6; SMACC response to the June Consultation, p.10.

coded terms to describe young children; an image of child for the chat icon; and links to CSAM in the chat description.[908]

12.12 Other civil society and industry stakeholders highlighted existing industry practice to identify and remove illegal content and commented on our proposed approach to inferring illegality. Snap, Inc. explained how it currently implements proactive technology to detect illegal content. This includes detection of group names that violate the providers policies.[909] Pinterest confirmed its own child sexual abuse policies that include images, videos, text and accounts, noting it has a "zero-tolerance" for illegal CSAM content on its service.[910] Similarly, the NSPCC said the proposal gave clarity on how to assess content for illegality, citing data indicating how services currently use metadata to detect suspicious behaviour.[911] The UK Association for Interactive Entertainment (Ukie) stated that its members agreed that tackling CSAM is a priority, stated that the gaming sector are already proactive in detection and removal of illegal material.[912] Some stakeholders recommended expanding the amendment to include additional coded terms and users' account and chat names that contain CSAM terms.[913]

12.13 While the majority of the responding stakeholders were supportive of our proposal, some expressed their disagreement with the amendments.[914] In these cases, the stakeholders raised issues such as the potential misuse of content reporting processes implicating innocent users; what constitutes "reasonably available information"; and providers making judgements on content with "limited evidence".[915] There was also a concern raised that the use of coded terms to detect CSAM may inadvertently capture content that is not illegal.[916]

12.14 We address the feedback in more detail throughout the 'our reasoning' section. However, some feedback did not directly relate to changes in our guidance and, as a result, we have included it in the relevant chapter in annex 1: Further stakeholder responses.

Our decision

12.15 After careful consideration, we have decided to broadly confirm the amendments to the ICJG for CSAM with some revisions. This means:

a) For scenario one, we are confirming our position without any changes.
b) For scenario two, we are confirming our position but revising it to narrow its application so that an illegal content judgement can be made where there are coded terms strongly

[908] IWF response to the June 2025 Consultation, pp.27-28.
[909] Snap, Inc. response to the June 2025 Consultation, pp.7-8.
[910] Pinterest response to the June 2025 Consultation, p.3.
[911] NSPCC response to the June 2025 Additional Safety Measures Consultation, p.32; O'Flaherty, K. (2021) All the data WhatsApp and Instagram send to Facebook. Wired.
[912] Ukie response to the June 2025 Additional Safety Measures Consultation, pp.22-23.
[913] C3P response to the June 2025 Consultation, pp.11-12; Lucy Faithfull Foundation response to the June 2025 Consultation, p.5.
[914] [✂]; Google LLC response to the June 2025 Consultation, p.19-20; [✂]; ICO response to the June 2025 Consultation, p.3; Luethje, Y. response to the June 2025 Consultation, p.4; Name Withheld 4 response to the June 2025 Consultation, p.5; [✂]; [✂]; [✂].
[915] Google LLC response to the June 2025 Consultation, p.19-20; ICO response to the June 2025 Consultation, p.3; [✂]; [✂]; [✂]; [✂].
[916] Name Withheld 4 response to the June 2025 Consultation, p.5; [✂].

associated with CSAM and a police report stating that CSAM has been identified, from a competent UK law enforcement authority.[917]

12.16 We have also decided to add a new scenario three which articulates that an illegal content judgement can be made where:

a) coded terms strongly associated with CSAM are present in the bio or heading,
b) there is a user complaint that includes evidence of CSAM, and
c) there is at least one of the following additional pieces of information from either of the following categories:[918]

i) account metadata relating to a user account's participation in other groups, chats or forums associated with the sharing or distribution of CSAM; or
ii) usernames, user profile images or user bio text containing a coded term or indicator strongly associated with CSAM.

12.17 Our guidance does not prevent a service provider from relying on other evidence to identify CSAM. Where it has robust evidence that a particular factor is strongly associated with CSAM and not with lawful activity, it should have regard to it.

12.18 In addition to these more substantive changes, we have made some minor revisions that include:

- Tightening the reference to "user complaints" as a form of other reasonably available information. We have removed this from scenario two, and moved this to scenario three, with the requirement that such complaints must be supported by compelling evidence.[919]
- Providing examples of what could be considered as reasonably available information for user-to-user services in our guidance.[920]
  - This includes for scenario three, identifying the types of other reasonably available information that providers should consider, including a specific type of account metadata and information contained in users' account names, user profile images or user bio text.
- Explaining what we mean by "coded terms strongly associated with CSAM" in our guidance.[921]

12.19 The full text of the amendment is set out in the amended ICJG, published alongside this statement. At the end of this chapter, we have provided a table giving the paragraph references.

---

[917] See paragraph 12.30 – 12.32 in this chapter for detail.
[918] We consider the reasonably available information specified in 12.16 c) (i) and (ii) to be indicative that messages, group chats or forums are being used to commit CSAM offences. While we recognise that the types of information specified are illustrative of child sexual abuse and exploitative behaviour, such as grooming, we consider that they may also indicate that the messages, group chat or forum is being used for the purposes of committing CSAM offences. We therefore consider this information, in combination with the other types of specified reasonably available information (i.e., a user complaint that includes evidence of CSAM and coded terms strongly associated with CSAM in the bio or heading) is sufficient to infer that the messages, group chat or forum is being used to facilitate the commission of CSAM offences.
[919] We explain what we mean by "compelling evidence" in paragraph 12.44.
[920] Ofcom (2026), Illegal Content Judgements Guidance (ICJG), paragraph 4.73 to 4.74, and 'Usage examples', pp.16-17.
[921] Ofcom (2026), ICJG, paragraph 4.70 to 4.71, p.16.

Our reasoning

12.20 Our amendments to the ICJG will strengthen service providers' ability to make illegal content judgements in relation to message threads, group chats or forums, in circumstances where it is not possible for them to review the underlying content. This may be the case where the underlying content is encrypted.

12.21 The amendments to the ICJG for CSAM set out three scenarios to offer guidance to service providers on:

i) inferring illegality based on content alone.
ii) inferring illegality based on coded terms strongly associated with CSAM in combination with a police report.
iii) inferring illegality based on coded terms strongly associated with CSAM in combination with a user complaint or report containing evidence of CSAM, and at least one further piece of reasonably available information (which is not a police report).

12.22 We set out each scenario in more detail below.

Scenario one: inferring illegality based on content alone

12.23 We indicated in the June 2025 Consultation that some content alone may be enough for providers to infer that other content within messages, group chats or forums is illegal CSAM content. This is the case even if the provider cannot review all the related content.

12.24 This scenario will apply when:

a) the image or icon for the messages, group chat or forum is CSAM;
b) the image or icon for the messages, group chat or forum is a still from a known CSAM video; or
c) the bio or heading for the messages, group chat or forum explicitly indicates that the contents are CSAM.

12.25 Our position remains unchanged from our June 2025 Consultation. We are not aware of any legitimate reason to use CSAM (whether an image or a CSAM URL), a still from a known CSAM video, or a bio or heading that explicitly states that the contents are CSAM, other than to signal to perpetrators that the purpose of the messages, group chat or forum is to commit one or more CSAM offences. In other words, the message thread, group chat or forum as a whole amounts to a conspiracy to commit one or more CSAM offences. Accordingly, this information alone should allow a provider to make a reasonable illegal content judgement in relation to the whole of the message thread, group chat or forum, unless there is clear evidence to suggest otherwise. We set out other examples in our ICJG to support providers' assessment of this information.

12.26 We consider that a provider will be able to identify some CSAM images on sight. However, it may be less obvious when the child is older or when the image concerned is a still from a known CSAM video. We expect some providers, to whom our recommendations on CSAM hash matching and URL detection apply, to have a database of such information available to them and as such that is "reasonably available information" for the purposes of the ICJG. Providers to which our recommendation to use hash matching does not apply may still become aware of the nature of such images through complaints or contact with law enforcement.

12.27 We do not expect this recommendation to add further resource burdens for service providers beyond those associated with their existing content moderation processes. Instead, it should assist providers in making judgements about content that is already detected. We consider the costs to providers of applying this guidance to be small and this amendment to be proportionate.

12.28 We did not receive evidence from our June 2025 Consultation that materially changed our assessment of this scenario. We therefore consider it an effective way to help reduce the presence of illegal CSAM online.

Scenario two: inferring illegality based on coded terms strongly associated with CSAM in combination with a police report

12.29 Scenario two pertains to cases where the bio or heading relating to a message thread, group chat or forum indicates that the content of the chat is CSAM through coded terms strongly associated with CSAM, in combination with a police report.

Police report

12.30 We consider a police report is a strong indicator for a provider to determine that specific content in a message, group chat or forum is CSAM. In this context, we use the term "police report" to mean a report from a competent UK law enforcement authority[922] identifying a specific message thread, group chat or forum and providing written confirmation that the authority has identified specific content within it, which in its opinion, amounts to CSAM.[923] [924]

12.31 We consider a police report allows service providers to rely on information from a competent UK law enforcement authority, reflecting its specialist expertise, intelligence and investigative information relating to CSAM offences. This provides a strong basis for assessing whether a particular message on a message thread, group chat or forum contains CSAM and may give providers greater confidence in their assessment of a piece of content's illegality. Where a service provider has received a police report, the information should be made available to its content moderation function for the purposes of moderating a relevant message thread, group chat or forum.

12.32 In our view, it is appropriate to place more weight on a police report in relation to the rest of a message thread, group chat or forum, than on a user complaint attaching evidence of CSAM. This is because of the risk that a user may deliberately post unlawful content in order to report it. However, we do not consider that a police report about a single message, on its own, is sufficient basis to infer that a message thread, group chat or forum as a whole amounts to a conspiracy to share CSAM. For this reason, we decided to recommend that service providers can only infer illegality from a police report, if coded terms strongly associated with CSAM are present in a bio or heading relating to a message, group chat or forum.

---

[922] For the purposes of this guidance, providers should treat a "police report" as a report from a competent UK law enforcement authority. Information from a non-UK law enforcement authority containing evidence of CSAM may still be relevant to an illegal content judgement, but it does not satisfy the police report requirement under this scenario.

[923] Generic statements that the use of particular coded terms is associated with CSAM or a general warning about child sexual exploitation and abuse (CSEA) offences unconnected to specific content or users will not be sufficient to constitute a police report for the purposes of this guidance.

[924] See Ofcom (2026), ICJG, paragraph 4.71, p.16.

Coded terms strongly associated with CSAM

12.33 Coded terms strongly associated with CSAM are phrases, keywords or other hints, including references to known CSAM series, which are commonly used by perpetrators to indicate or signpost the presence of CSAM within message threads, group chats or forums without stating this explicitly. A provider may identify coded terms through its own moderation processes, and/or reliable evidence sources, such as the IWF Keywords list, or information from law enforcement agencies.[925] This information should be accessible to the content moderation function to support its assessment of the message threads, group chats or forums identified in the police report.

12.34 We consider the presence of such coded terms when combined with a police report, indicates a high likelihood that the purpose of the relevant message thread, group chat or forum is to commit CSAM offences. We consider it is reasonable for a service provider to take appropriate moderation action, including removal of the relevant message thread, group chat or forum.

12.35 We do not expect the review of the reasonably available information for scenario two (i.e., the bio or heading related to a message thread, group chat or forum) to add further resource burdens for service providers beyond those associated to their existing content moderation processes. Instead, it should assist providers in making content judgements of content that is already detected. We consider that the costs for service providers applying this guidance to be small and this amendment to be proportionate.

Stakeholder feedback related to scenario two

12.36 We received stakeholder feedback that resulted in minor changes to scenario two.

Coded terms strongly associated with CSAM

12.37 We received feedback that queried the detection of CSAM through coded terms. This included whether providers' detection of CSAM through coded terms could inadvertently capture content that is not illegal CSAM.[926] Ukie stated that applying measures to coded terms could inadvertently capture content from other communities that use coded language (such as fandom groups or competitive teams), potentially undermining trust in providers' moderation processes.[927] Name Withheld 4 also cited the example of a service provider trying to prevent discussion of CSAM by banning anyone using the term 'cp', which can be associated with the term "child pornography" in some contexts. It suggested this led to many users being falsely accused.[928]

12.38 We appreciate that terms, such as 'cp', may be used in many other contexts unrelated to CSAM, and used in legitimate and unrelated contexts. Relying on such a term in isolation could lead to incorrect judgements or the inappropriate sanctioning of users. Because of this, we do not consider that the use of a single coded term in a bio or heading relating to a message, group or forum, on its own, would provide reasonable grounds to infer that the content within amounts to an offence. However, if such a coded term is combined with other pieces of information, such as a police report (scenario two) we consider this would be a sufficient basis to infer illegality in a message, group chat or forum. In our view, the presence of the illegal content provides reasonable grounds to infer that the code is not

[925] Ofcom (2026), ICJG, paragraph 4.70 to 4.71.
[926] Name withheld 4 responses to the June 2025 Consultation, p.5; [✂].
[927] Ukie response to the June 2025 Consultation, p.23.
[928] Name Withheld 4 response to the June 2025 Consultation, p.5.

being used innocently, and that the purpose of the message thread, group chat or forum is to commit CSAM offences.

Adversarial use of indicators and content reporting processes

12.39 Providers asked that the guidance allow for procedural safeguards against false positives, noting the risk that indicators may be "adversarially manipulated".[929]

12.40 We recognise that some users may make adversarial use of CSAM and content reporting processes to implicate innocent users. This may include hacking, changing group settings and false reporting.[930] A malicious actor may, for example, apply a coded term to a message thread, group chat or forum bio and then report it to the provider with the intention of triggering an incorrect illegal content judgement.

12.41 We also recognise that, given the risk of adversarial misuse, a coded term combined with a user complaint may not in some cases provide a sufficient basis for an illegal content judgement. We are concerned that the outcome of such an incorrect assessment would cause a significant interference with a user's freedom of expression.

12.42 Taking this into consideration, we decided to revise scenario two by removing the recommendation to consider either a user complaint or police report. Rather, the scenario now recommends that police reports are the only form of reasonably available information that may be used on its own, alongside coded terms strongly associated with CSAM. This change recognises stakeholder feedback regarding the adversarial use of the content reporting process, and the risk of false complaints potentially leading to incorrect illegal content judgements, wrongful removal of content and interference with users' freedom of expression.[931] We consider that, where coded terms are present, this revision allows service providers to rely on information from a competent UK law enforcement authority, reflecting its specialist expertise, intelligence and investigative information relating to CSAM offences. We also expect the revision to reduce false complaints and avoid wrongly accusing innocent users.

Scenario three: inferring illegality based on coded terms strongly associated with CSAM, a user complaint, and at least one other piece of information (that is not a police report)

12.43 We have decided to include a new scenario to provide greater clarity on the other circumstances in which service providers should infer illegality where they cannot review the underlying content.

12.44 Under this scenario, service providers should infer that the content within the messages, group chat or forum is CSAM where:

a) coded terms strongly associated with CSAM are present in the bio or heading; and
b) there is a user complaint that includes evidence of CSAM; and

---

[929] [✂].
[930] See the section 'Rights impact' for a further discussion of malicious reporting and freedom of expression considerations.
[931] Under the Act, service providers are required to take appropriate action on complaints by a user if the provider has given a warning to the user, suspended or banned the user from 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.

c) there is at least one of the following additional pieces of reasonably available information from either of the following categories:

i) account metadata relating to a user account's participation in other message groups, chats or forums associated with the sharing or distribution of CSAM; or
ii) usernames, user profile images or user bio text containing a coded term or indicator strongly associated with CSAM.

12.45 Our guidance does not prevent a service provider from relying on other evidence to identify CSAM. Where it has robust evidence that a particular factor is strongly associated with CSAM and not with lawful activity, it should have regard to it.

12.46 We do not expect the review of the reasonably available information for scenario three (i.e., account metadata or users' account names, user profile images, or user bio text containing coded terms associated with CSAM) will add significant resource burdens in addition to the potential impacts set out as part of a provider's content moderation processes. Instead, it should assist providers in making content judgements of content that is already detected. We consider that the costs of applying this guidance to be small and this amendment to be proportionate.

12.47 We provide more information on our rationale for recommending each type of reasonably available information in the following paragraphs.

Coded terms strongly associated with CSAM

12.48 We have already set out more detail on our rationale as to why coded terms are useful but not sufficient on their own for identifying CSAM in message threads, chats and forums in paragraph 12.33 to 12.35.

User complaint that includes evidence of CSAM

12.49 Our decision to include this scenario is based on our position that complaints handling is feasible where service providers are unable to view end-to-end encrypted content. For example, if a user submits a complaint to an end-to-end encrypted service provider, it may automatically take a copy of the (unencrypted) content concerned and the content immediately around it and attach such content to the complaint.[932] Where such a complaint is made and the copied content is CSAM, we regard it as strong evidence that the content being complained about is CSAM. We do not consider that a user complaint which does not include such a copy is evidence of CSAM, due to the prevalence of malicious reporting. The user complaint should be made available to content moderators to support their assessment of the relevant messages, group chats or forums.

12.50 While we consider a user complaint with a copy of content to be strong evidence that the item of content concerned is CSAM, our position is that additional information is still important to adequately assess the illegality of content of the whole message thread, group chat or forum (that is not technically feasible to review). This is because a user could post CSAM themselves, with a view to reporting it, and because the presence of one item of CSAM does not, on its own, show that the message thread, group chat or forum as a whole amounts to a conspiracy to commit CSAM offences.

12.51 For this reason, we decided that service providers should consider coded terms strongly associated with CSAM in a bio or header and user complaint containing evidence of CSAM in combination with at least one additional specified piece of reasonably available information,

[932] Ofcom (2024), December 2024 Statement, Volume 2, paragraph 6.36.

such as account metadata or coded terms strongly associated with CSAM in users' account names, user profile images or user bio text.

Account metadata

12.52 Service providers should, in addition, consider account metadata relating to a user account's participation in other groups, chats or forums associated with the sharing or distribution of CSAM. We consider evidence that a user is involved in other chats which share CSAM corroborates an inference that the chat in question is a conspiracy to share CSAM.

Usernames, user profile images or user bio text containing a coded term or indicator strongly associated with CSAM

12.53 Users' account names, user profile images or user bio text can include coded terms or indicators strongly associated with CSAM, such as victims' or survivors' names, titles of known CSAM series or non-CSAM stills from known CSAM videos. This type of material could indicate to service providers that an account is associated with CSAM offending. Where a user with these characteristics is using a message thread, group chat or forum whose bio or heading contains coded terms strongly associated with CSAM and in relation to which a user complaint shows CSAM to have been present, we consider this to be a strong basis on which to infer that the purpose of the chat is to commit CSAM offences. We detail the types of coded terms associated with CSAM that may be considered reasonably available information in paragraph 12.33.

12.54 However, users' account names containing CSAM related terms may also be ambiguous and not consistently an indicator of CSAM being present. In particular, we consider the use of victims' or survivors' names could produce a significant number of false positives for service providers, which presents a risk to users' rights. There will be legitimate reasons why victims and survivors use their own names when they participate online. For this reason, we recommend that service providers only consider victims' and survivors' names as reasonably available information if those names are included in recognised established keyword lists compiled by trusted organisations, such as the Internet Watch Foundation, which identify words, phrases and codes used to conceal CSAM online.

12.55 We also consider that non-CSAM stills taken from known CSAM videos may be used as user profile images. We consider that if a member of a message thread, group chat or forum has a profile image that is a non-CSAM still taken from a known CSAM video that is a possible indicator that CSAM may be present in the group. Where these images can be identified through trusted sources, such as CSAM hash databases, we consider that this should be considered as reasonably available information.

12.56 We also recognise that providers' ability to view or review users' account names, user profile images and user bio text varies depending on the type of service. We therefore consider that this information is reasonably available where it is technically feasible for a service provider to review users' account names, user profile images and user bio text of the users on the message thread, chat or forum in question.

Stakeholder feedback resulting in new scenario

12.57 In the following paragraphs, we detail the stakeholder feedback that led to the inclusion of this new scenario.

"Reasonably available information"

12.58 We received feedback from an industry stakeholder requesting clarity on what constitutes "reasonably available information" in our guidance.[933] It suggested that our guidance may be disproportionate if providers are encouraged to actively seek out this information.[934]

12.59 To address this concern, we added a third scenario to clarify the types of reasonably available information that content moderators should use when making an illegal content judgement. We set out examples of the types of reasonably available information (not including police reports) that could be present alongside coded terms when inferring illegality.[935] We consider the prescribed reasonably available information appropriately reflects available information to a service provider for the purposes of making an illegal content judgement.

12.60 However, we recognise that service providers may, in practice, have access to further information beyond what is specified in this guidance which may be relevant to illegal content judgements. In these cases, our guidance does not prevent a service provider from taking into account any additional information, if it is processed lawfully, in accordance with data protection legislation.[936]

12.61 To note, service providers may not need to consider all available information in every case, where it is possible to make an accurate illegal content judgement on the basis of more limited information.

Coded terms strongly associated with CSAM

12.62 We received feedback on the use of coded terms strongly associated with CSAM, which is set out under scenario two (paragraphs 12.33 to 12.35). However, this feedback was also considered in the development of scenario three.

12.63 We decided that, where a coded term is combined with a user complaint containing evidence of CSAM and at least one other specified piece of reasonably available information, it provides a sufficient basis for a provider to infer the illegality of a message thread, group chat or forum.

Limited evidence to make judgements

12.64 We received feedback that raised concerns about service providers making judgements on content with limited evidence.[937] Hammy Media, while separately making it clear it has "a zero tolerance policy towards illegal content, especially CSAM" raised concerns with relying on user bio or heading information to infer that content should be CSAM or CSAM discussion, suggesting that no clear criteria had been provided to identify such content in practice.[938] Similarly, the ICO noted a risk of inaccurate illegal content judgements based on names, icons or bios "without consideration of the substantive content of the group".[939]

---

[933] Google LLC response to the June 2025 Consultation, pp.19-20.
[934] Google LLC response to the June 2025 Consultation, p.9.
[935] See paragraph 12.44 in this chapter.
[936] See the section 'Rights impact' for a discussion of related privacy and data protection issues.
[937] Google LLC response to the June 2025 Consultation, p.9; [✂]; ICO response to the June 2025 Consultation, p.3; [✂]; [✂].
[938] Hammy Media Ltd response to the June 2025 Consultation, p.5; Hammy Media Ltd email dated 8 October 2026.
[939] ICO response to the June 2025 Consultation, p.3.

12.65 Google LLC also stated that there is a risk providers incorrectly impose sanctions and additional consequences (such as police reporting), on the basis of potentially very limited evidence, especially where "broader evidence is not reasonably available and there is no evidence to the contrary".[940] Snap, Inc. stated (in relation to scenario one) that it may not be appropriate to automatically remove an entire 1:1 or group chat based on a single instance of CSAM posted in the chat. It suggested that recipients of the illegal content "may not have intentionally sought out or engaged with the CSAM content" and "have no control over what other users send to them".[941]

12.66 We set out our reasons for judging that it is reasonable to infer that a chat, forum or messaging group is for sharing CSAM in our assessments of each scenario.

User account names, profile images and bio text containing CSAM related indicators

12.67 We received feedback from the Lucy Faithfull Foundation proposing an expansion to the amendment to include users who create account names that contain CSAM terms. It felt this would act as a preventative step before harm is caused, and suggested providers supplying a deterrence message at this point to discourage further offending.[942]

12.68 We agree that users' account names, user profile images or user bio text containing CSAM related terms, including victims or survivors' names, known CSAM series or non-CSAM stills from known CSAM videos, could indicate to service providers that an account is associated to a perpetrator of CSAM and, therefore, has a higher likelihood of the content itself being CSAM. However, we do not consider that CSAM coded terms or indicators in users' account names, user profile images or user bio text should be relied on in isolation. This information may support an illegal content judgement where they are considered alongside coded terms strongly associated with CSAM in the bio or heading and a user complaint containing evidence of CSAM. We consider that requiring this combination of additional information reduces the risk of an incorrect judgement and makes reliance on these indicators reasonable.

12.69 Taking this all into consideration, we designed scenario three to include coded terms strongly associated with CSAM in users' account names, profile images, and bio text, as reasonably available information that service providers should consider when inferring illegality of content. We set out our rationale for including users' account names, user profile images, and user bio text that contain coded terms associated with CSAM as a category of reasonably available information in the explanation of scenario three in paragraphs 12.53 to 12.56.

Additional contextual information

12.70 We received feedback in response to our position that perpetrators of CSAM may use phrases, keywords or other hints to indicate or signpost the presence of CSAM content within messages, group chats or forums.[943] Specifically, we indicated that while these phrases may not themselves be illegal, when combined with other information referred to in paragraph 12.44 (captured under scenario three), this may provide a sufficient basis for a

[940] Google LLC response to the June 2025 Consultation, p.9.
[941] Snap, Inc. response to the June 2025 Consultation, p.8.
[942] Lucy Faithfull Foundation response to the June 2025 Consultation, p.5.
[943] Ofcom (2025), June 2025 Additional Safety Measures Consultation, p.123.

provider to reasonably infer that the content of the messages, chat group or forum are CSAM.[944]

12.71 Stakeholders recommended additional types of contextual information which may indicate the presence of CSAM in a message, group chat or forum.[945] Pinterest indicated that reviewing specific off-content signals and metadata relating to group chats for signposting or "breadcrumbing" may be helpful for services with large group messaging functionalities. However, it further indicated that it may be more effective and proportionate to allocate time and resources to prioritise reviewing off-content signals relating to account metadata for service providers that do not offer large group messaging functionality.[946] The Canadian Centre for Child Protection (C3P) added that services could also reasonably assess "accounts that add high volumes of users to a group, while sharing no common network; accounts that have excessively high outgoing message to incoming message ratios; [and] accounts that send high volumes of messages to users not found in their personal contacts, etc."[947]

12.72 We do not consider that a broad recommendation to use "account metadata" or guidance referring to a "high volume of messages" alone provides sufficient grounds to infer that the content in messages, group chats or forums amounts to a relevant offence. Many people send high volumes of messages for lawful reasons, particularly commercial reasons. However, where a provider has robust evidence that a particular factor is strongly associated with CSAM and not with lawful activity, it should have regard to it.

Rights assessment

Introduction

12.73 In this section, we consider the impacts of this new guidance on users' rights under Article 8 (right to respect for private and family life), Article 10 (right to freedom of expression) and Article 11 (freedom of association) of the European Convention on Human Rights (ECHR). Any interference with these rights must be in accordance with[948] or prescribed by[949] law and necessary in a democratic society in pursuit of a legitimate aim including prevention of crime or disorder, protection of health or morals and the protection of the rights and freedoms of others. In either case, to be necessary, the restriction must correspond to a pressing social need, and it must be proportionate to the legitimate aim pursued.

12.74 We recognise the approach set out above may amount to a significant interference with users' rights to freedom of expression and privacy. Freedom of association is also of particular relevance because the removal of content may affect users' ability to communicate and associate with others, including where lawful content is affected by an illegal content judgement.

12.75 Balanced against these rights we also acknowledge that the guidance contributes to removing and preventing CSAM content online. CSAM content is a particularly serious harm, and its prevalence and circulation online are widespread, and technology is increasingly complex, reshaping crime itself, accelerating it, globalising it and making it more harmful

[944] Ofcom (2025), June 2025 Additional Safety Measures Consultation, p.125-126.
[945] Canadian Centre for Child Protection (C3P) response to the June 2025 Consultation, pp.11-12.; Pinterest response to the June 2025 Consultation, p.4.
[946] Pinterest response to the June 2025 Consultation, pp.3-4.
[947] C3P response to the June 2025 Consultation, pp.11-12.
[948] Article 8 of the European Convention on Human Rights (ECHR).
[949] Article 8 of the ECHR.

causes serious and long-term harm to children, and significantly interferes with the rights of victims and survivors.[950]

12.76 As part of our assessment, we have also considered stakeholder feedback which raised concerns about the impact of our proposals on users' rights.[951] Google LLC and another stakeholder raised concerns that the guidance may lead to disproportionate interferences with users' freedom of expression, particularly where providers are required to infer the presence of CSAM without seeing it, on the basis of limited evidence which may be unreliable, prone to error and with a risk that it is taken out of context.[952] Concerns were also raised by a stakeholder that reliance on signals such as user bio or heading information could lead to errors including the misinterpretation and the removal of lawful content.[953]

12.77 We address stakeholder feedback regarding inferring the presence of CSAM without seeing the content (or based on limited evidence) in paragraphs 12.99 to 12.103. We also recognise that there is a greater potential for an illegal content judgement to be incorrect where content is taken down without reviewing it and that this could have a significant impact on rights because it affects all the content on the message thread, chat or forum. We have therefore considered the risk of over-removal, the possibility of error where providers infer illegality without direct review of content, and the need for sufficiently robust safeguards in our assessment below.

12.78 We explain below why, in the defined circumstances addressed by the guidance, we consider that any resulting interference with users' rights is necessary and proportionate.

Freedom of expression and association

Potential interference with freedom of expression and association

12.79 We are giving guidance to providers to assist them in their decisions as to whether to take down content in a message thread, chat or forum that they have not directly reviewed on the basis that there are reasonable grounds to infer that the thread, chat or forum as a whole amounts a conspiracy to commit one or more CSAM offences. We recognise that this approach is a significant and unusual regulatory step to take. Although the Act requires the removal of illegal content, we acknowledge that there are very significant impacts on freedom of expression and association if lawful content is wrongfully removed, as this may affect users' ability to communicate and associate with others, and disrupt legitimate and lawful interactions in an online group. It may also lead to user sanctions, including a user being permanently banned from a service and prevented from returning.[954] As set out in section 66 of the Act, providers are also required to report CSEA content of which they are aware to the NCA, (unless it has been appropriately reported to equivalent overseas bodies).

---

[950] NCA (2026). 'NCA launches the National Strategic Assessment 2026'.
[951] Google LLC response to the June 2025 Consultation, p.9; [✂]; Name Withheld 4 response to the June 2025 Consultation, p.5; [✂]; Porter, J. response to the June 2025 Consultation, pp.3-4; [✂]; [✂]; [✂].
[952] Google LLC response to the June 2025 Consultation, p.9; [✂].
[953] Hammy Media Ltd response to the June 2025 Consultation, p.5.
[954] We note, however, our user banning recommendation provides that users on a message thread, chat or forum who have not themselves shared, generated, or uploaded CSEA to the service should not be subject to a ban. Those who have received CSAM and not viewed it should not be subject to a ban, and providers may choose not to ban those who have received and viewed CSAM but appropriately reported and deleted the content. This reduces the risk of unnecessary interference with their rights to freedom of expression or association.

Where, following assessment of the provider's report, the NCA investigates, this can have serious negative consequences on users' wellbeing, livelihood and reputation.

12.80 These risks of interference with users' rights arise because, in the defined scenarios addressed by the guidance, providers make an inference about the content in a message thread, chat or forum without being able to review all of its contents directly. The approach may therefore result in the removal of lawful content, including lawful content posted by users who are not themselves engaged in unlawful activity.

12.81 In assessing the level of interference, we note that an illegal content judgement for the purposes of the guidance is not a criminal finding against any individual user. It does not give rise to a criminal record or to criminal sanctions. However, we acknowledge that the interference with affected users' rights is serious, particularly if a lawful chat, forum or message thread is mistakenly identified as illegal.

12.82 Taking those matters together, we consider that the approach may amount to significant interference with users' rights to freedom of expression and freedom of association.

Prescribed by law

12.83 The interference is prescribed by law – the UK Parliament has designated CSAM as priority illegal content under the Act and has imposed duties on providers intended to secure its swift detection and removal. Section 193 of the Act requires Ofcom to produce guidance about illegal content judgements for the purposes of the illegal content safety duties, including the takedown duty and the risk assessment duty. This guidance is issued pursuant to that statutory framework and explains the limited circumstances in which providers may infer that the contents of a message thread, chat or forum is to commission CSAM offences.

12.84 The guidance confines the approach to defined scenarios and identifies the evidential basis required in each case. In that way, it is intended to be adequately accessible, certain and foreseeable for providers. It also sits within a broader statutory framework directed at protecting children from serious harm, consistent with the state's positive obligations to protect children from child sexual abuse and exploitation.

12.85 We recognise that the approach may, in some cases, result in lawful content being erroneously removed, and that scenarios two and three require providers to assess coded terms and other contextual indicators which may be context sensitive. Whilst we recommend that providers use established keyword lists compiled by organisations such as the Internet Watch Foundation which identify words, phrases and codes used to conceal CSAM online, we do not consider it appropriate to prescribe an exhaustive list of such indicators in guidance, as their meaning may change over time and perpetrators may adapt their language and behaviour to avoid detection. In addition, the guidance limits the circumstances in which such indicators may be relied on. We are not guiding providers to treat coded terms as sufficiently probative in isolation, but only to consider them in combination with other indicators in the defined scenarios set out above. In our view, our approach is sufficiently clear for providers to apply and formulated with sufficient precision to enable users to foresee the circumstances in which it may be applied and the likely consequences that may follow.

Necessary for a legitimate aim

12.86 The objectives of the approach correspond to the legitimate aims under Articles 10 and 11, including the prevention of crime, the protection of public morals and the protection of the rights and freedoms of others. We know that perpetrators use private messaging to circulate

CSAM in order to evade detection, therefore the guidance addresses this offending and seeks to prevent the creation, distribution and continued prevalence and circulation of CSAM online in private groups, to prevent users from inadvertently encountering CSAM in them and to protect the rights of others, in particular children who are victims and survivors of child sexual abuse.

12.87 We consider that the approach is necessary - and no more than the minimum necessary - in a democratic society. It addresses a pressing social need to prevent the creation, distribution and continued circulation of CSAM in private groups. Limiting action to content that can be directly reviewed would risk CSAM remaining in circulation while law enforcement authorities separately seek to identify and take action against offenders. In practice, such action can be challenging as offenders operate individually or in networks across multiple jurisdictions and change or conceal their identities to evade detection. As a result, harmful content may remain online and the associated harm to children continue.

12.88 We have considered whether a less intrusive approach could achieve the objective. We have considered whether providers could be limited to removing only content they are able to review directly, or whether action by providers should be deferred pending investigation and action by law enforcement agencies. We do not however consider that those alternatives would be effective in the circumstances addressed by the guidance. Removing only directly reviewable content may leave CSAM within the wider thread, chat or forum. The broader removal of the message thread, chat or forum is therefore necessary to address the harm effectively. Deferring action until law enforcement authorities can identify offenders and take action may permit continued circulation of CSAM for a potentially significant period. In this regard, we do not consider that a less intrusive measure would achieve the objective without unacceptably compromising its effectiveness.

12.89 We therefore consider it necessary, in the limited and clearly defined circumstances we have set out, to give guidance to providers to support their ability to make illegal content judgements in these circumstances by reference to the reasonably available information that they have available to them (and to ensure that the approach taken is sufficiently robust given the risks we have identified).

Proportionality of the interference

12.90 We consider that the objective pursued by the guidance is sufficiently important to justify a significant interference with users' rights. It is directed at preventing the creation, distribution and continued circulation of CSAM and at protecting children, including children depicted in that material, from grave and ongoing harm.

12.91 The approach is rationally connected to the objective of preventing the creation, distribution and continued circulation of CSAM. It is directed at circumstances where there is reasonably available evidence which indicates that the purpose of the relevant message thread, chat or forum is to commit CSAM offences, even though the provider cannot review every item of content within it. Where it is reasonable to infer from available information that the thread, chat or forum as a whole amounts to an ongoing conspiracy to commit CSAM offences, it is reasonable to infer that every message on it is illegal content as a part of that conspiracy. Removal of the wider thread, chat or forum is capable of interrupting continued circulation of CSAM in a way that removing only directly reviewable content would not. The three scenarios support that connection by identifying the circumstances in which that inference can reasonably be made.

12.92 We have carefully considered whether the reasonably available information in each scenario is sufficiently probative to give providers reasonable grounds to infer illegality.

12.93 In scenario one, we consider that it is reasonable to infer that the relevant message thread, chat or forum is for sharing CSAM if the image or icon for the chat is CSAM, a still from a known CSAM video or it otherwise explicitly indicates that the contents are CSAM because we are not aware of any reason for using these explicit indicators other than to signal to other perpetrators that the purpose of the chat is to commit one or more further CSAM offences.

12.94 In scenarios two and three, the indicator in a bio or heading is a coded term strongly associated with CSAM, rather than a CSAM image, still, or explicit CSAM reference. We acknowledge stakeholder feedback about the risk of incorrect judgements based on limited evidence. We have therefore limited these scenarios to circumstances where a coded term is considered together with specified categories of reasonably available information, reducing the risk of incorrect judgements and limiting any resulting interference with users' rights.

12.95 Following our consideration of feedback, we consider that our narrowed scenario two, requiring a coded term plus a police report from a competent UK law enforcement authority identifying a specific message thread, group chat or forum and providing written confirmation that the authority has identified specific content within it, which, in their opinion, amounts to CSAM, provides sufficient evidence to give providers reasonable grounds on which to infer the presence of CSAM. Requiring corroboration from a police report reduces the risk of interference with rights and targets the approach to strongly probative indicators.

12.96 In scenario three, providers may infer illegality only where coded terms strongly associated with CSAM are supported by a user complaint containing evidence that the content complained about is CSAM, together with a further corroborating indicator. We explain why we consider a user complaint to be effective in paragraphs 12.49 to 12.51. We recognise that a coded term alongside a user complaint may not always provide a sufficiently reliable basis for removing an entire message thread, chat or forum. However, we consider that the presence of a coded term, in combination with a user complaint containing compelling evidence of CSAM plus an additional indicator reduces the risk of an incorrect judgement to near zero, and provides a sufficiently reliable basis to make an inference of illegality.

12.97 We acknowledge that scenarios involving coded terms carry a greater risk of interference with users' rights than scenario one. However, with the corroboration required by multiple sources of information in scenarios two and three, we consider that any resulting interference is justified and proportionate in light of the seriousness of the harm, the limits of the scenarios, and the further safeguards discussed in paragraphs 12.99 to 12.103.

12.98 We recognise that the approach may remove content, which considered in isolation may appear to be lawful content, alongside clearly illegal content. Where innocuous content has been posted by users in the course of an ongoing conspiracy to commit CSAM offences, we consider it to be a part of the conspiracy. Even if that were incorrect, we would consider that consequence to be justified by the need to put an end to the forum in which the offending is taking place. We also acknowledge stakeholder feedback about the impact of the guidance on lawful users and lawful messages within a thread, chat or forum being used for an illegal purpose without their knowledge. However, because the guidance is limited to circumstances where a provider has clear evidence that the very purpose of the message

thread, chat or forum is to commit CSAM offences, we consider our approach to be proportionate and limits the impact on users' freedom of expression and association.

12.99 Finally, across all three scenarios, the approach is supported by safeguards designed to mitigate the risk of incorrect judgements. In particular, user complaints must contain sufficiently precise and adequately substantiated information to demonstrate that the relevant message thread, chat or forum contains CSAM. Police reports must contain confirmation that the law enforcement authority has identified specific content within it, which, in their opinion, amounts to CSAM.

12.100 User complaints and appeals mechanisms are a further safeguard. Our reporting and complaints measure ICU D10 of the Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes) allows users to appeal against a ban and have it overturned if the provider determines that the relevant content was not illegal content, thereby ending the wrongful interference with their rights.[955] These mechanisms are particularly important where the approach may affect users whose own content was lawful or who were not themselves engaged in unlawful activity.

12.101 We acknowledge that these safeguards would not necessarily remedy the whole of the harm experienced by the user (in particular, it may take some time for the appeal to be dealt with); nor do they eliminate the possibility that lawful content may be removed or of a provider making an incorrect judgement. They do, however, materially reduce the likelihood and severity of unjustified interference.

12.102 We consider that data protection law provides further safeguards. When processing users' personal data providers must comply with applicable data protection legislation. ICO has issued substantial guidance to help them do this. The principle of accuracy is among the principles at the heart of providers' processing of personal data. The requirement to ensure personal data is accurate incentivises providers to, for example, take an appropriate and evidenced approach to identifying coded terms relating to CSAM, reducing the risk of lawful content being incorrectly identified as illegal content and of lawful users being subject to investigation by law enforcement. If a provider chooses to use automated processing to make illegal content judgements, data protection law provides further safeguards, including as to appeals.

12.103 We therefore consider that providers' data protection obligations including the requirement to process personal data accurately, provide an additional safeguard against unjustified interference with users' rights to freedom of expression.

12.104 Taking these matters into account, and in light of the grave harm caused to children by child sexual abuse, the interference with the rights of victims and survivors caused by CSAM and its continued circulation, we consider that the guidance strikes a fair balance between the seriousness of any interference with users' rights to freedom of expression and freedom of association and the importance of the objective pursued. We therefore consider that any interference with those rights is proportionate.

Privacy

12.105 We have also considered the impact of the approach on users' right to respect for private life and correspondence under Article 8. In particular, it may result in the removal of private communications within a message thread, chat or forum and require providers to process

[955] Illegal Content Codes, Measure ICU D10.

information associated with the users on the suspect chats, forums and messaging threads.[956] This may include information dating back over an extended period of time where it remains available to the provider.

12.106 It may also involve providers taking further action in relation to an incorrect illegal content judgement, such as reporting users to law enforcement agencies which would involve the collection, use and disclosure of personal data. We also acknowledge that the approach may involve a provider considering not only information associated with a user suspected of involvement in CSAM offences, but also that associated with other users participating in a message thread, chat or forum, who may be unaware of its purpose.

12.107 Where an illegal content judgement is incorrect, the interference with users' rights may be significant. In those circumstances, lawful private communications may be removed, inaccurate information may be generated about users' involvement in a thread, chat or forum associated with CSAM, and a user may be subject to sanctions or reported to law enforcement on an incorrect basis. Those outcomes may affect users' private life, correspondence, reputation and wellbeing, including where they were not themselves engaged in unlawful activity.

12.108 Some stakeholders raised concerns about privacy impacts stating that the proposals would harm privacy and encryption.[957] techUK, which also separately stated that it supports the objectives of the measure, highlighted the need for regulatory clarity on how the proposals interacted with data protection law, so that services can be confident of meeting both sets of obligations.[958] The ICO requested clarification and safeguards in our guidance to ensure providers' decisions are accurate and fair under data protection law, especially due to the potential impacts of a user being classified as having shared CSAM.[959] A provider asked that the guidance allow for procedural safeguards against false positives, noting the risk that indicators may be adversarially manipulated.[960]

12.109 We recognise stakeholders' concerns and have made changes to our initial proposals to minimise these risks. In paragraphs 12.115 to 12.123 we discuss the safeguards that reduce the risk of unlawful interference with users' privacy rights in relation to our approach.

In accordance with the law

12.110 The interference is in accordance with the law. For the reasons set out above, the approach forms part of the statutory framework established by the Act, including the illegal content safety duties and Ofcom's duty under section 193 to produce guidance on illegal content judgements. The guidance confines the approach to defined scenarios and identifies the evidential basis required in each case, so that the circumstances in which providers may make illegal content judgements are sufficiently clear and foreseeable.

Necessary for a legitimate aim

12.111 The interference pursues legitimate aims under Article 8. It is directed at the prevention of crime and the protection of the rights and freedoms of others, in particular children who are

---

[956] The metadata and contextual information that a provider may consider is set out in paragraphs 12.52 to 12.60.
[957] ICO response to the June 2025 Consultation, p.19; [✂]; Pinterest response to the June 2025 Consultation, p.4; [✂]; Thompson, A. response to the June 2025 Consultation, p.4.
[958] techUK response to the June 2025 Consultation, p.13, techUK email dated 12 October 2026.
[959] ICO response to the June 2025 Consultation, p.19.
[960] [✂].

victims and survivors of CSEA. Preventing the creation, distribution and continued circulation of CSAM is a particularly important objective.

12.112 We consider that the interference with users' privacy is necessary in a democratic society. There is a pressing social need to prevent perpetrators from exploiting message threads, chats or forums in circumstances where it is not technically feasible for providers to review all content, including where perpetrators seek to use coded language and indirect references to CSAM. If providers were limited to removing only the content they could individually review, there would be a risk that CSAM would remain available within the wider message thread, chat or forum.

12.113 We have considered whether a less intrusive approach could achieve the objective. In particular, we have considered whether providers could be limited to removing only content they are able to review directly, or whether action should be deferred pending investigation and action by law enforcement agencies. We do not however consider that those alternatives would be effective in the circumstances addressed by the guidance because that may lead to CSAM remaining in circulation in a thread, chat or forum for a potentially significant period.

12.114 We have also considered whether the guidance should exclude certain categories of reasonably available information, in particular contextual information in users' account names and profiles from the information that providers may take into account. However, we do not consider that approach would be effective because such information may, in some circumstances, provide important context which supports the identification of CSAM, and excluding it could result in CSAM remaining in circulation for a potentially significant period.

Proportionality of the interference

12.115 We recognise that the connection between the approach set out in the guidance and the legitimate aim of preventing and disrupting CSAM offending is weaker where intrusive action is taken on the basis of ambiguous or less probative indicators alone. However, we consider that the approach is rationally connected to that aim because the guidance includes safeguards which limit the approach to circumstances where providers may only rely on explicit indicators, or on a sufficiently probative combination of coded indicators, alongside police reports or a combination of other contextual indicators including user complaints and reports, users' participation in other groups, chats or forums associated with the sharing of CSAM, and users' account names to make an illegal content judgement. These safeguards will assist providers to identify and remove CSAM from their platforms and prevent its continued circulation online.

12.116 The proposed approach seeks to reduce unnecessary interference with privacy rights by limiting reliance on ambiguous or weak indications alone, and in the circumstances of scenarios two and three requires a combination of indicators. We recognise that coded terms in a bio alongside a user complaint may be vulnerable to malicious misuse and that treating this combination of indicators as being sufficient for a provider to rely on to reasonably infer illegality would present a greater risk of disproportionate interference with users' privacy rights.

12.117 In developing our approach, we have incorporated a number of safeguards and identified wider safeguards under the Act. Taken together these safeguards mitigate against a range of risks, including in particular the risk of an illegal content judgement being made on the basis of unreliable indicators.

12.118 We consider that the risk of unjustified interference with privacy is mitigated because the guidance is confined to limited circumstances where there is clear evidence, or a sufficiently robust combination of indicators that the relevant message thread, chat or forum is being used for the commission or facilitation of CSAM offences. We recognise that there are greater privacy and data protection concerns in relation to providers considering coded terms and other indicators, as these involve further processing of personal data. We have introduced safeguards intended to reduce the risk of error, including the requirement for combinations of indicators to be considered, clarification of what constitutes a police report, and what must be included in a user complaint. These safeguards reduce the risk of inaccurate inferences and unfair processing in relation to identified or identifiable users.

12.119 We recognise that a user account's participation in other message groups, chats or forums associated with the sharing or distribution of CSAM or other contextual indicators such as users' account names, user profile images or user bio text containing coded terms strongly associated with CSAM may in some cases be ambiguous, be capable of innocent explanation or insufficiently reliable in isolation. We consider however that these indicators may be probative where multiple indicators corroborate one another and/or where they are used in combination with coded terms or a reliable user report which contains evidence of CSAM offending.

12.120 We further note that, under the Act, a user can appeal if they believe a provider has taken action against them or their content is removed based on an incorrect illegal content judgement and that a provider will take steps to rectify inaccurate personal data. This acts as an additional and important safeguard.

12.121 We also recognise the risk that an incorrect illegal content judgement could result in users being reported to the NCA, involving the disclosure of information about the user and potentially other identifiable individuals resulting in significant risks to users' right to privacy. We recognise that a report to the NCA based on an incorrect illegal content judgement would be a significant intrusion into an individual's privacy, even where NCA checking or triage processes result in no further action being taken. However, we consider that the approach set out in this guidance appropriately mitigates this risk by limiting the circumstances in which providers are required to make illegal content judgements and requiring reliance on sufficiently probative indicators. These safeguards reduce the likelihood of incorrect reporting and ensure that any interference with privacy rights is limited to what is necessary and proportionate in pursuit of the legitimate aim of preventing and detecting CSAM offences.

12.122 We also note that, when processing users' personal data, providers must comply with applicable data protection legislation and have regard to relevant guidance from the ICO. This includes applying appropriate safeguards to protect users' rights, including requirements relating to transparency, fairness, data minimisation and accuracy. Those requirements are particularly relevant where providers process personal data for the purposes of making an illegal content judgement, including where a provider creates personal data about a user by inferring illegality. The principle of accuracy is also a particularly important safeguard where a provider creates or processes personal data in order to report identified CSEA content to the NCA or another law enforcement agency.

12.123 Nothing in our Codes or guidance, as far as it relates to the processing of personal data, either supersedes or derogates from the requirements of data protection legislation. The UK GDPR restricts the circumstances in which decisions can be made based solely on automated

processing of personal data, where the decision has legal or similarly significant effects for the relevant individual.

12.124 We also recognise that the removal of CSAM acts directly to protect the rights of victims and survivors depicted in CSAM. This sort of content causes significant and ongoing harm to victims and survivors from knowing that the material continues to circulate online (or in some cases themselves viewing that material), or from being identified by persons who have viewed that material. Its removal protects victims' and survivors' privacy rights and protects their personal data.

12.125 We recognise that the approach may interfere significantly with users' privacy, in particular where lawful private communications are removed or where an incorrect illegal content judgement results in inaccurate inferences about users or disclosure of information about a user to law enforcement. We are mindful that a fair balance must be struck between our starting point that CSAM is a serious harm, the interference with victims and survivors rights that it causes and the interference with users' privacy rights.

12.126 Taking these matters together, including the safeguards identified above, we consider the approach strikes an appropriate balance between the interference with users' privacy rights and the need to prevent the creation, distribution and continued circulation of CSAM, and to protect the rights of children and victims and survivors of child sexual abuse.

Conclusion

12.127 There is evidence to demonstrate that perpetrators use messaging, group chats or forums to facilitate the commission of CSAM offences without detection. This is particularly the case in respect of offences of distribution, or encouragement to distribute illegal images of children, but several other CSAM offences are also committed in this way. The June 2025 Consultation feedback supported this conclusion. However, we recognise stakeholder concerns that highlighted areas where we could better support providers' assessment of this content. This resulted in a change to the original amendment, to help providers' assessment of this content. See paragraphs 12.15 to 12.19 for further detail.

12.128 We consider that the impact of our guidance on users' rights to freedom of expression and privacy is likely to be proportionate relative to the significant benefits of reducing the prevalence and distribution of CSAM online, particularly in circumstances where it is not technically feasible for providers to review the content in messages, group chats or forums.

12.129 We are therefore confirming our position that in the circumstances set out in the three scenarios, providers may infer from the specified reasonably available information that messages, group chats and forums contain CSAM and are proceeding with the amendments as outlined in paragraphs 12.15 to 12.19.

12.130 The amendment is included in our Illegal Content Judgements Guidance (ICJG), published alongside this statement. The following table gives the paragraph references.

12.131 We have also corrected a number of minor errors within the guidance document more widely.

Table 12.1: Amendments to the ICJG

Previous version	Amendments	Notes
Paragraphs 4.1 – 4.10	Paragraphs 4.1 – 4.10	No change
N/A	Paragraph 4.11	New paragraph added
Paragraph 4.11 – 4.61	Paragraphs 4.12 – 4.62	No change
N/A	Paragraphs 4.63 – 4.85	New paragraphs added (including, after paragraph 4.85, new grey boxes titled 'Usage examples' and 'Reasonably available information for user-to-user services')

13. Combined Impact Assessment

What is this chapter about?

In the preceding chapters of this statement, we assess the individual impact of each of the decisions we have taken in terms of new Code measures or amendments to the Codes. In this chapter, we assess the combined impact of the measures as a package, including their expected impact on economic growth. Having considered the combined impact on different groups of services and on economic growth, we explain why we have determined that the package of measures is proportionate.

Introduction

13.1 In the preceding chapters, we have assessed the impacts of each of the decisions we have taken to add or modify measures in our Illegal Content and Protection of Children Codes of Practice for user-to-user services (together referred to as Codes) individually and concluded that each is proportionate. In this chapter, we consider the combined impact of these decisions overall by looking at the following factors:

- **Whether each measure has distinct benefits to contribute to the overall package of measures.**[961] [962] This analysis is an important part of our assessment of whether, taken together, our decisions are proportionate. If the measures were substantially duplicative of one another, the aggregate impact of our decisions could be disproportionate even if, looked at in isolation, each individual measure delivered net benefits. As part of this assessment, we have also considered the impact of 'existing measures' in the Codes, alongside the 'new' measures described in this document to ensure the existing and new measures do not overlap in such a way as to render measures redundant.
- **Whether the overall impact on services is proportionate, particularly for smaller services.** In assessing this, we recognise that the costs of the whole package may be significant for some service providers, even when the cost of the individual measures may not be significant.

13.2 In this chapter we assess the combined impact for user-to-user services in scope of the Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes), followed

[961] The measures for using hash matching to detect image-based intimate image abuse content for user-to-user and search services (ICU 14 and ICS C8 respectively) were recommended in our May 2026 Statement on Detecting intimate image abuse and are therefore no longer in this package of measures. Similarly, the measures for crisis response (ICU C15, ICU C16, PCU C11, PCU C12) were recommended in our June 2026 Statement on Crisis response protocol and is therefore no longer in this package of measures. We assessed the combined impact of those measures in those statements.

[962] We are now recommending one measure for proactive technology for unknown image-based CSAM; further details can be found in chapter 9: Deployment of proactive technology to detect image-based CSAM.

by the combined impact on user-to-user services in scope of the Protection of Children Code of Practice for user-to-user services (Protection of Children Code).[963]

Growth duty

13.3 Ofcom is required, when exercising our regulatory functions, to have regard to the desirability of promoting economic growth, including through considering the importance of ensuring the regulatory action we take is necessary and proportionate.[964] This is referred to as the 'growth duty'.

13.4 We are also required to have regard to the UK Government's statutory guidance on the growth duty.[965] This guidance explains that the duty needs to be considered alongside our other statutory duties and that its purpose is not to achieve or pursue economic growth at the expense of necessary protections.[966] Among other things, it also identifies particular drivers of economic growth, including innovation, investment and competition.

13.5 The growth duty has applied to our online safety functions since 6 April 2026 following the end of a time-limited exclusion for these functions. As such, we did not expressly consider the growth duty in our June 2025 Additional Safety Measures Consultation (June 2025 Consultation). For this statement, we have assessed both the impact of the measures and the combined impact, by considering the wider economic impacts of our decisions.

Illegal Content user-to-user Codes

13.6 In our June 2025 Consultation, we considered the impact of the overall package of measures that we proposed for providers of different kinds of services in scope of the Illegal Content user-to-user Codes. We provisionally concluded that the proposed measures were proportionate as they targeted specific harms and/or functionalities with clear links to harm, which our assessment showed were not adequately addressed by existing measures.

13.7 We received a range of views from stakeholders in response to our June 2025 Consultation Combined Impact Assessment:

- Some stakeholders agreed our additional package of measures are proportionate.[967] Some said the proposed measures struck a proportionate balance between reducing serious harms and recognising the different sizes and types of services.[968] Others

[963] The intimate image abuse hash matching measure (ICS C8) was the only measure in the June 2025 Consultation which applied to search services, and the impact of this on search services was assessed as part of our May 2026 Statement. Since consultation we have added a content moderation measure setting out the steps services should take to reverse incorrect decisions, which is recommended for all user-to-user and search services. This measure is not expected to have a significant additional impact on search services, for reasons set out in chapter 9: Deployment of proactive technology to detect image-based CSAM. The remainder of this combined impact assessment is therefore only focused on user-to-user services.
[964] Section 108 of the Deregulation Act 2015.
[965] Section 110(3) of the Deregulation Act 2015.
[966] Department for Business and Trade, 2024. Growth Duty: Statutory Guidance – Refresh.
[967] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.25; Check My Ads Institute response to the June 2025 Consultation, pp.15-16; Digital Resilience in Education response to the June 2025 Consultation, p.17; Evangelical Alliance UK response to the June 2025 Consultation, p.13; [✂]; Social Media Animal Cruelty Coalition (SMACC) response to the June 2025 Consultation, p.17.
[968] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.25.

suggested the current system "is not effective or efficient" and described the new measures as a "positive move in the right direction".[969]

- Several stakeholders considered our additional package of measures for smaller services disproportionate.[970] They suggested that the proposed measures could place a disproportionate burden on smaller or low risk providers and create a negative impact on competition in the market.
- Other stakeholders suggested that our additional package of measures are disproportionate.[971] Some were specifically concerned about the impact on specialist services.[972] Twitch suggested that the measures were disproportionate for services that already maintain high safety and innovation standards.[973] Stakeholders also raised concerns about the impact of the measures on privacy and freedom of expression.[974] [975] [976]

Our decision and reasoning

13.8 Having considered the stakeholder feedback, we have concluded that the combined impact of the package is proportionate for the following reasons:

- Each individual measure delivers distinct benefits;
- we have taken a risk-based approach and have recommended only one additional measure for small low risk services;
- where we have imposed significant costs on smaller services, this is justified by their risk profile; and
- the overall cost for large services is proportionate.

13.9 We expand on these points in the following section.

Each individual measure delivers distinct benefits

13.10 In this statement we recommend additional measures for user-to-user services in scope of the Illegal Content Codes to further improve online safety beyond existing measures following our December 2024 Statement on Protecting People from Illegal Harms Online (December 2024 Statement). For the list of measures see Tables 13.1 and 13.2 later in this chapter.

---

[969] National Police Chiefs' Council (NPCC) response to the June 2025 Consultation, pp.18-19.
[970] ACT – The App Association response to the June 2025 Consultation, p.12; [✂]; [✂]; [✂].
[971] ACT – The App Association response to the June 2025 Consultation, p.12; Name Withheld 9 response to the June 2025 Consultation, p.13; [✂]; [✂]; GitHub response to the June 2025 Consultation, p.11; Harris, B. response to the June 2025 Consultation, p.10; Luethje, Y. response to the June 2025 Consultation, p.10; Name Withheld 4 response to the June 2025 Consultation, p.11; [✂]; [✂]; [✂]; [✂]; Twitch response to the June 2025 Consultation pp.22-23; [✂].
[972] ACT – The App Association response to the June 2025 Consultation, p.12; Name Withheld 4 response to the June 2025 Consultation, p.10; GitHub response to the June 2025 Consultation, p.11; [✂]; [✂]; [✂].
[973] Twitch response to the June 2025 Consultation, pp.22-23.
[974] ACT – The App Association response to the June 2025 Consultation, p.13; [✂]; Harris, B. response to the June 2025 Consultation, p.10; [✂].
[975] Privacy concerns are addressed in relevant chapters and in annex 3: Legal framework.
[976] Freedom of expression concerns are addressed in relevant chapters, chapter 14: Statutory tests and in annex 3: Legal framework and annex 2: Equality impact assessment and Welsh Language Assessment.

13.11 We consider that the benefits of the measures do not substantially overlap with each other, or with existing measures. Most of the measures target specific severe harms and/or functionalities with clear link to harm, which our assessment shows are not adequately addressed by existing measures.

13.12 Where more than one measure targets the same harm, they focus on different aspects of it and deliver distinct benefits:

- Some measures address the same harm, but in distinct and complementary ways. For example, two of our proactive technology measures and our user banning measure target the distribution of child sexual exploitation and abuse (CSEA) material. Proactive technology is key to detecting and removing child sexual abuse material (CSAM), enabling providers to take appropriate action. Hash matching is effective at matching known CSAM, and other classes of technology are capable of identifying CSAM, including CSAM that has not been previously identified.[977] User banning aims to remove access to services from users who generate, share or receive this material. We consider that together these measures will help reduce current CSAM content and reduce its future upload. As CSEA harm can manifest in different ways, more than one measure is required to address it.
- Certain measures in this statement strengthen those already included in the Illegal Content Codes. For example, we previously recommended safety defaults for child user accounts.[978] These measures target particular functionalities,[979] restrict visibility and engagement between child users and other users, and give child users relevant supportive information at critical points in a child users journey. We have strengthened these measures by recommending the application of the measures to all child users, either through the use of highly effective age assurance or by applying the safety defaults to all users on the service.

13.13 Overall, our assessment shows that each measure has distinct benefits. The measures we have decided to add to our Codes in this statement build on existing measures, and in many cases they are complementary. There is not significant duplication between the measures.

We recommend two additional measures for all user-to-user services, including smaller low risk services

13.14 In our June 2025 Consultation, we proposed very limited additional measures for smaller services that are low risk. We considered that including multiple costly measures for such services was unlikely to be proportionate to the benefits they could provide in terms of reducing relevant harms.

13.15 Following further assessment of the costs of – and feedback on – the user sanctions policy measure (measure ICU H2) we are no longer recommending this measure for smaller, low

[977] See the chapter 9: Deployment of proactive technology to detect image-based CSAM for a description of the relationship between our proactive technology measure to detect unknown image-based CSAM content and our hash matching measures.
[978] Illegal Content Codes of Practice for user-to-user services (Illegal Content Codes), Measures ICU F1 and ICU F2.
[979] Illegal Content Codes, Measure ICU F1 targets functionalities which have been identified as risk factors for grooming harm on user-to-user services. These include network expansion prompts, direct messaging, connection lists and automated information displays.

risk services.[980] We are therefore only including two additional measures for such services, outlined in Table 13.1.

Table 13.1: Additional measure in the Illegal Content user-to-user Codes for all user-to-user services

Measure	Short description	Who this measure applies to
User banning and preventing return following detection of CSEA (ICU H3)	Services should ban users who share, generate, or upload CSEA, and take steps to prevent their return to the service for the duration of the ban	Providers of all user-to-user services
Steps to take on reversal of content decisions (ICU C4A)	Services should reverse actions where a provider identifies that it has incorrectly taken action on content or a user	Providers of all user-to-user services

13.16 We consider the cost of the user banning measure to be proportionate for smaller low risk user-to-user services, even when this is added to the cost of other existing measures that apply to them. As explained in chapter 3: User sanctions, we expect costs to be relatively low for smaller low risk services and there is flexibility for services to apply these measures in a cost-effective way. Given the severity of the harm, and the regular movement of CSEA perpetrators between different services, we also consider that if a user is found to have posted CSAM or engaged in grooming it is important that they are banned irrespective of the size and risk profile of the service. We consider that having a measure setting out that incorrect action on content or a user should be reversed is a critical part of protecting users' rights to freedom of expression, and that the costs of this measure will generally be limited to content identified in a service's quality assurance activities.

The overall cost for smaller user-to-user services with specific risks and/or functionalities is proportionate

13.17 Some of the measures in the Illegal Content Codes apply to services that have specific risks and/or functionalities. These are summarised in Table 13.2 and largely apply to relevant user-to-user services regardless of size.[981] To ensure proportionality, some of the measures apply to smaller services only where they assessed as high risk of a relevant harm, whereas they apply to large services at both medium and high risk levels. This approach reflects both resource constraints and the greater potential impact of harmful content on larger services.

13.18 Following the June 2025 Consultation, we have decided to no longer recommend two of the measures originally proposed for one-to-many livestreaming services.[982]

[980] See a full explanation in chapter 3: User sanctions.
[981] In addition to the measures in Table 13.2, we recommend some amendments to the existing measures in the Illegal Content user-to-user Codes as outlined in the relevant chapters. We consider the incremental cost of these amendments to be relatively low and consider them proportionate.
[982] In the June 2025 Consultation we proposed to recommend that one-to-many livestreaming services should restrict certain functionalities when children were livestreaming to help prevent various forms of harm, and

Table 13.2: Additional measures in the Illegal Content Codes for user-to-user services with relevant risks or functionalities

Measure	Short description	Who this measure applies to
Recommender systems (ICU E2, ICU E3)	Providers should design and operate their recommender systems to ensure that content indicated as potentially certain kinds of priority illegal content (relevant illegal content) is excluded from the recommender feeds of users not determined to be adults by highly effective age assurance.	Providers of user-to-user services likely to be accessed by children that have a content recommender system, and are medium or high risk for: a) hate and/or terrorism b) suicide and self-harm.
Livestreaming (ICU D17)	The provider should have a mechanism to enable users to report that a livestream contains content that depicts the risk of imminent physical harm.	Providers of user-to-user services offering one-to-many livestreaming that meet either of the following: are medium or high risk for one or more relevant harms.[983]
User-to-user safety defaults (ICU F1, ICU F2)	Amendments to ICU F1 and F2. These changes mean that services in scope of ICU F1 and F2 should implement these measures using one of the following approaches: • using highly effective age assurance, or • applying ICU F1 and F2 to all users of the service.	Providers of user-to-user services that are high risk for grooming; and Providers of large user-to-user services that are medium risk for grooming.
Highly effective age assurance in the Illegal Content User-to-User Codes (ICU B1, ICU D15, ICU D16)	Defining highly effective age assurance for the purposes of the Illegal Content User-to-User Codes and setting out principles providers should have regard to when implementing highly effective age assurance. Providers should allow for appeals of highly effective age assurance decisions and take appropriate action when these are upheld.	Providers of user-to-user services that use an age assurance process to determine which UK users of the service are child users.

have human content moderators available whenever users were able to livestream. Our decision and reasoning are outlined in chapter 2: Livestreaming.

[983] The relevant harms are terrorism content, grooming or image-based child sexual abuse material (CSAM), assisting or encouraging suicide and self-harm, hate, harassment, stalking, threats and abuse offences, or animal cruelty. This includes services that only provide livestreaming and those where it is an ancillary feature.

Measure	Short description	Who this measure applies to
Hash matching for terror content (ICU13)	Providers use perceptual hash matching to detect terrorism content so it can be removed	Providers of the following user-to-user services that enable regulated user-generated content[984] to be generated, uploaded or shared: • Large services that are high or medium risk for terrorism content • Services that are high risk for terrorism content and have more than 700,000 monthly active UK users • File-sharing/file-storage services that are high risk for terrorism content
User sanctions (ICU H2)	Providers should put in place a sanctions policy in relation to UK users who generate, upload, or share illegal content and/or illegal content proxy with the objective of preventing future dissemination of relevant illegal content.	Providers of user-to-user services that are either large and/or multi-risk in relation to illegal content.[985]
Proactive technology (ICU PT1)	Providers should use proactive technology to detect unknown image-based CSAM as it manifests on their service.	Providers of user-to-user services that are: • Large services that are high or medium risk for image-based CSAM • Services that are high risk for image-based CSAM and have more than 700,000 monthly active UK users • File-sharing/file-storage services that are high risk for image-based CSAM

13.19 We acknowledge that the overall package of existing measures for the services in scope of the additional measures in Table 13.2 could entail substantial costs.

[984] User-generated content in the form of photographs, videos, or visual images (whether or not combined with written material).
[985] A service is multi-risk if it is medium or high risk for two or more illegal harms.

13.20 The additional measures set out in Table 13.2 only apply to small services where there is a significant risk of harm and/or risk of displacement. The services in scope are those that pose a significant risk and substantial harm to users, regardless of their size.

13.21 Without these measures, the risks of harm to users would not be adequately addressed.

13.22 Some smaller service providers, including small specialist service providers, may struggle to resource the measures. However, even if this were to happen, we would still consider the package proportionate due to the severity of the harm being addressed, the distinct benefits of each of the measures in reducing these risks, and the fact that we are targeting the measures at higher risk services. We have not seen sufficient evidence to suggest otherwise, and we consider the costs are justified by the significant additional benefits that arise from the measures individually and collectively. We explain this further in paragraphs 13.36 - 13.42.

13.23 Several measures in this statement recommend that services use highly effective age assurance to determine if users are adults and we consider this to be proportionate. Each measure using highly effective age assurance will result in additional costs for providers, as more users may seek to verify themselves as adults in order to access an expanding set of features. Service providers will also need to adapt their user experience to work effectively with highly effective age assurance. However, we consider that where service providers that have already implemented highly effective age assurance for existing measures, a large part of the associated cost will have already been incurred, and this cost will be spread over several measures.

13.24 We provide an alternative option for highly effective age assurance in relation to the user-to-user safety defaults measures and the recommender systems measures. Providers may either (i) apply measures ICU F1 and F2, and ICU E2 and E3, to all users, or (ii) use highly effective age assurance to apply these measures only to users who are not determined to be adults. This gives providers flexibility to implement the measures in a way that is most appropriate for their service.

13.25 Some of the measures in this statement are likely to increase the cost of implementing existing measures. For example, some of the measures (such as proactive technologies, including hash matching) aim to improve a service's ability to identify relevant illegal content, which could increase the cost of implementing existing content moderation measures. We consider this proportionate as the measures target services with specific harms or functionalities where we consider the risk to be significant. The volume and cost of content moderation should increase with the size and risk assessment of the service, and therefore also result in greater benefits.

13.26 Some providers may already have well developed safety systems in place. Under the Act, the measures that Ofcom includes in its Codes are recommendations for how the services can comply with their statutory safety duties. Providers may retain their existing approach where they consider that their safety features already comply with a relevant duty. Where providers use alternative measures, they must have particular regard to the importance of protecting UK users' right to freedom of expression and privacy and must keep a record of the measures taken and explain how those measures meet the relevant duties.[986]

[986] Full explanation in chapter 1: Introduction, our duties and navigating the statement.

The overall cost for large user-to-user services is proportionate

13.27 We recognise that large service providers could incur significant cost to implement existing measures and the recommended additional measures in Tables 13.1 and 13.2 may increase this further. However, we expect providers of large services to have the resources to undertake these measures. Even if some service providers have more limited resources, the increase in overall cost would be proportionate given the reduction in risk which could impact a significant number of users on these services.

Protection of Children user-to-user Code

13.28 In our June 2025 Consultation, we considered the impact of the overall package of additions to our Protection of Children Code. We provisionally concluded that the proposed measures targeted specific harms and/or functionalities with clear links to harm, which our assessment showed were not adequately addressed by existing measures.

13.29 We received a range of views from stakeholders in response to our June 2025 Consultation Combined Impact Assessment:

- Some stakeholders suggested that the additional measures for smaller user-to-user services in scope of the Protection of Children Code were disproportionate.[987] Some stakeholders raised concerns about the impact on competition and suggested too many of the measures applied to services regardless of size or risk.[988]
- Some stakeholders agreed that our additional package of measures are proportionate, describing them as offering strong protections and representing a significant step towards ensuring children's safety online.[989]
- Several stakeholders considered our additional package of measures to be disproportionate.[990] Some argued that the measures were overly burdensome and disproportionate, citing the closure of UK hobby websites as evidence.[991] Stakeholders also raised concerns about the impact these measures have on privacy and freedom of expression, suggesting these measures may have a negative impact on encrypted services and that highly effective age assurance reduced freedom of expression.[992]

Our decision and reasoning

13.30 Having considered the stakeholder feedback, we have concluded that the combined impact of the package is proportionate for the following reasons:

---

[987] ACT – The App Association response to the June 2025 Consultation, p.12; [✂]; [✂]; [✂].
[988] ACT – The App Association response to the June 2025 Consultation, p.12; [✂]; [✂].
[989] A member of the Philippine Survivor Network response to the June 2025 Consultation, p.26; Check My Ads Institute response to the June 2025 Consultation, p.16; Digital Resilience in Education response to the June 2025 Consultation, p.18; Evangelical Alliance UK response to the June 2025 Consultation, p.13; [✂]; NPCC response to the June 2025 Consultation, p.18; SMACC response to the June 2025 Consultation, pp.17-19.
[990] ACT – The App Association response to the June 2025 Consultation, p.12; Name Withheld 9 response to the June 2025 Consultation, p.13; [✂]; [✂]; Luethje, Y. response to the June 2025 Consultation, p.10.; Name Withheld 4 response to the June 2025 Consultation, p.10; Snap, Inc. response to the June 2025 Consultation, p.20; [✂]; [✂].
[991] Name Withheld 4 response to the June 2025 Consultation, p.10.
[992] Freedom of expression concerns are addressed in relevant chapters, chapter 14: Statutory tests and in annex 2: Equality impact assessment and Welsh Language Assessment and annex 3: Legal framework.

- the additional measure delivers distinct benefits;
- we have taken a risk-based approach and included no additional measures for small low risk user-to-user services;
- the overall cost for large user-to-user services is proportionate.

13.31 Following feedback on the user sanctions policy measure (PCU H2), we are no longer recommending it for all user-to-user services likely to be accessed by children that prohibit at least one specific type of primary priority content, priority content, and/or non-designated content. Instead, the measure is now recommended only for services that meet those criteria that are large and/or multi-risk (see chapter 3: User sanctions).

13.32 As a result of this change, and the change to our proactive technology measure, we are now only recommending two additions to our Protection of Children user-to-user Code.[993] We are only recommending one addition for small, low risk user-to-user services. We consider that the change to the user sanctions measure removes any additional cost for small low risk services and responds directly to stakeholder feedback suggesting that the previous package was too much of a burden for small, low risk services, including UK hobby sites.

13.33 In Table 13.3, we outline two additional measures in the Protection of Children Code recommended to providers of user-to-user services.

Table 13.3: Additional measure in the Protection of Children Code for user-to-user services

Measure	Short description	Who this measure applies to
User sanctions (PCU H2)	Services should put in place a sanctions policy in respect of UK users who generate, upload, or share content harmful to children and/or harmful content proxy, with the objective of preventing future dissemination of relevant content that is harmful to children.	Providers of user-to-user services likely to be accessed by children that: • prohibit at least one specific type of primary priority content, priority content and/or non-designated content; and • are either large and/or multi-risk in relation to content harmful to children.
Steps to take on reversal of content decisions (PCU C4A)	Services should reverse actions where a provider identifies that is has incorrectly taken action on content or a user	Providers of all user-to-user services

13.34 The user sanctions measure in table 13.3 aims to improve child safety beyond existing measures included in our April 2025 Statement on Protecting Children from Harms Online (April 2025 Statement). We consider it has distinct benefits that do not substantially overlap with our existing measures for the same reasons set out earlier for the equivalent illegal harm measures (see paragraphs 13.20-13.25). We expect the cost of this measure will vary

[993] We will consider the impact of proactive technology measures including their impact on existing Codes measures in a future consultation.

widely across service providers, increasing with service size and risk level.[994] We consider the overall cost of this measure proportionate, when accounting for existing measures, for the same reasons set out in the equivalent illegal content measures. We have not seen sufficient evidence to suggest this assessment should differ.

13.35 We consider that having a measure setting out that incorrect action on content or a user should be reversed is a critical part of protecting users' rights to freedom of expression, and that the costs of this measure will generally be limited to content identified in a service's quality assurance activities.

We consider the overall package proportionate in terms of its expected impact on economic growth

13.36 We consider the additional measures to the Illegal Content and Protection of Children Codes to be both necessary and proportionate when assessed against the Growth Duty. In line with the statutory guidance, we have considered the combined impact of these measures on key drivers of economic growth, including innovation, investment and competition.[995] As outlined, these measures are targeted at services with risks of specific harms or that have functionalities that pose particular risks. We therefore judge the impact on growth to be proportionate considering the nature of the risks being addressed.

13.37 We have taken steps to minimise any negative impact on economic growth. By recommending very few additional measures for smaller services, we consider we have limited any increase in costs for services attempting to enter the market. This approach recognises that smaller providers often face proportionally higher regulatory compliance costs and therefore aims to avoid any unnecessary costs on service providers with more limited resources. It also avoids imposing unnecessary costs where the expected additional safety benefits are limited.

13.38 We recognise that, under this approach, some providers may face higher costs as their services expand. However, we consider this approach a proportionate consequence of our risk based approach. As the service grows, the number of users who may benefit from stronger safety measures also increases, as does the scale of harm if risks are not addressed. It is also reasonable to expect that, as a service grows their capacity to implement additional safety measures will generally increase. We therefore consider it appropriate that additional measures apply where the expected safety benefits justify the additional costs.

13.39 We consider that there are elements of the additional safety measures that may have positive impacts on growth. For example, measures (ICU PT1) recommend that providers use proactive technology to detect unknown image-based CSAM. This could result in an increased demand for proactive technologies and incentivise investment and innovation in these areas for businesses that specialise in automated content detection tools.

13.40 Services have significant flexibility in their adoption of the measures, and we expect that they will take steps to minimise their costs. For example, some measures give services options in how they apply them, such as the recommender systems measures and the user-to-user safety defaults measures. Service providers can either apply these measures to all

[994] See chapter 3: User sanctions, for the full costs and impacts of measure PCU H2.
[995] Department for Business and Trade, 2024. Growth Duty: Statutory Guidance – Refresh.

the users on the service or use highly effective age assurance to apply them only to users not determined to be adults, whichever they consider the most proportionate for their service.

13.41 In meeting the outcomes set out in the Codes, providers of regulated services may adopt innovative technical or organisational measures that are effective in addressing the relevant risks. Where possible, the measures take a technology-neutral approach and are not intended to mandate the use of specific technologies, system architectures, or model types. Service providers may achieve the required outcomes through alternative measures, allowing them the flexibility to implement the recommended measures in a proportionate way that minimises unnecessary burden.[996] We do not expect any significant adverse impact on adjacent markets or the wider economy from these measures.

13.42 Given the significant benefits the measures we are recommending will have, the fact that we have taken a risk based approach and sought to avoid imposing burdensome requirements on small low risk services, and the flexibility we have afforded service providers, we consider that any impact on growth from our decisions is likely to be limited and proportionate.

Conclusion

13.43 Our assessment shows that the package of measures is proportionate and contributes to reducing the risks online of illegal harms and improving safety for children online. Each measure delivers a distinct benefit which goes beyond, and is additional to, both the other measures in this package and those of the measures covered in our previous statements. While the overall cost could be significant for some services, our assessment for the purposes of this statement is that this is proportionate given the risks of harm to users and the incremental benefit of each measure in the package in reducing these risks.

13.44 We conclude that the package of measures is justified and proportionate based on the assessments we have set out for each measure in earlier sections of this chapter. In consideration of our responsibilities under the Deregulation Act 2015, we also consider the package of measures proportionate relative to their impact on growth.

[996] Service providers do not need to follow the Codes and may seek to comply with their safety duties by taking what the Online Safety Act 2023 (the Act) calls “alternative measures”. Where providers take alternative measures, they must keep a record of what they have done and explain how the relevant safety duties have been met. They should also have particular regard to the importance of the rights of their users to freedom of expression within the law and protecting the privacy of users.

14.Statutory Tests

What is this chapter about?

In designing the Codes, the Online Safety Act 2023 (the Act) requires Ofcom to have regard to a number of principles and objectives, set out in Schedule 4 to the Act. The Communications Act 2003 (the 2003 Act) also places a number of duties on Ofcom in carrying out our functions.

In this chapter, we set out the matters to which we must have regard under the Act and the 2003 Act, and explain the reasons why we think the recommendations in the Codes meet them.

Background

14.1 In designing the codes of practice, the Online Safety Act 2023 (the 'Act') requires Ofcom to have regard to a number of principles and objectives, set out in Schedule 4 to the Act. The Communications Act 2003 (the '2003 Act') also places a number of duties on Ofcom in carrying out our functions, including requiring Ofcom to have regard to the risk of harm to citizens presented by content on regulated services.

14.2 In Chapters 2-11, we set out the recommended Codes measures. An overview of these measures can be found in 'Summary of our decisions', and our combined impact assessment of the codes' measures can be found in chapter 13: Combined impact assessment. The complete measures can be found in the 'Illegal Content Code of Practice for User-to-user services', the 'Illegal Content Code of Practice for User-to-User Services: Proactive Technologies (unknown content)', the Illegal Content Code of Practice for Search Services, the Protection of Children Code of Practice for user-to-user services and the Protection of Children Code of Practice for search services. We provide further information regarding Ofcom's duties relating to the preparation of the Codes in chapter 1: Introduction, our duties and navigating the statement.

14.3 We consider that the Codes measures meet the requirements set out in Schedule 4 to the Act and section 3 of the 2003 Act. In this chapter, we take each of the relevant requirements in turn and set out how we have met them.

14.4 In response to the June 2025 Consultation, we received stakeholder feedback relevant to our duties under section 3 of the 2003 Act and Schedule 4 to the Act. We consider and respond to this feedback in our analysis of individual measures set out in the relevant chapters in this statement and in annex 1: Further stakeholder responses. Below we set out an overview of how the recommended measures fulfil the requirements of section 3 of the 2003 Act and Schedule 4 to the Act.

14.5 We have already published our decisions relating to two proposals made in the June 2025 Consultation. In May 2026, we published our statement on Detecting Intimate Image Abuse (May 2026 Statement).[997] In June 2026, we published our statement on Crisis Response Protocols (June 2026 Statement).[998] We direct you to those statements for an overview of

[997] Ofcom (2026), Statement Detecting Intimate Image Abuse, published 18 May 2026.
[998] Ofcom (2026), Statement Crisis response protocol, published 09 June 2026.

how those recommended measures fulfil the requirements of section 3 of the 2003 Act and Schedule 4 to the Act.

Duties and principles

The Communications Act 2003

14.6 As required by section 3 of the 2003 Act, in making the recommendations in the Codes, Ofcom has had regard to the matters set out below and to the risk of harm to citizens presented by content on regulated services.

> Section 3(1): It shall be the principal duty of Ofcom, in carrying out their functions: a) to further the interests of citizens in relation to communication matters; and b) to further the interests of consumers in relevant markets, where appropriate by promoting competition.

14.7 We have set out in this statement how recommended Codes measures will mitigate the risks of illegal harm to all users, and the risks of harm to children online, thereby furthering their interests as well as the interests of citizens in the UK more generally.

14.8 We have considered the interests of consumers in relevant markets (particularly users of regulated services) as part of our assessment of the proportionality of our recommendations, including any potential impacts on the provision of services to users.

14.9 We have considered the impact of our measures on adult and child users, including in light of our approach to age assurance, meaning that certain measures may apply to all users where a service is not using highly effective age assurance.

14.10 We have also considered the rights of users and other interested persons in our rights assessment for each measure, where we consider any impacts of the measure on users' (children, and adults' where relevant) rights, including their rights to freedom of expression, data protection and privacy, as required by the Act.

> Section 3(3): In performing their duties under subsection (1), Ofcom must have regard in all cases to (a) the principles under which regulatory activities should be transparent, accountable, proportionate, consistent and targeted only at cases in which action is needed, and (b) any other principles appearing to Ofcom to represent best regulatory practice.

14.11 In the interest of transparency, accountability and fairness (and as required by the Act) [999], we consulted stakeholders on our proposals, and our consultation includes impact assessments for each of the measures we propose to include in the Codes. We are setting out clearly the evidence and assumptions used to arrive at our proposals. In the relevant chapters we explain how we approach impact assessments.

14.12 Our impact assessments of measures consider effectiveness, costs, rights, and other relevant factors and explain why we consider the measures are proportionate to the benefits to children and adults as relevant. We consider the proportionality of the package of the

[999] Under section 3(3) of the 2003 Act, Ofcom must, in the performance of their duties under subsection (1), have regard to the principles under which regulatory activities should be transparent, accountable, proportionate, consistent and targeted (only at cases in which action is needed. Ofcom must also have regard to any other principles appearing to us to represent best regulatory practice. This includes the public law duty to act fairly.

measures as a whole in chapter 13: Combined impact assessment. See our impact assessment guidance for more information on how we approach impact assessments.[1000]

14.13 The measures are informed by our assessment of the risk posed by illegal content to all users and the risks of harm to children. We have prioritised developing proposed measures that can effectively mitigate the significant risks identified in our analysis and those required by the Act and have targeted our proposed measures at the kinds of services which we think should be deploying them because this would lead to the greatest benefits, given the risks they pose.

14.14 We consider the measures in the Codes to be a proportionate means of supporting services to meet the illegal content safety duties and children's safety duties under the Act. Our reasoning for our final decisions about the Codes is set out in this statement, together with the evidence we have relied on in reaching our decisions, contributing to transparency and accountability.

> Section 3(2)(g): In carrying out our functions, Ofcom are required to secure the adequate protection of citizens from harm presented by content on regulated services, through the appropriate use by providers of such services of systems and processes designed to reduce the risk of such harm.

14.15 The changes made to our Codes of Practice are intended to mitigate the risks to users from illegal content, and the risks to children from content harmful to them. For example:

- Our measures on the use of perceptual hash matching to detect terrorism content and CSAM is designed to reduce the amount of illegal content that reaches UK users;
- Our recommendations around content recommender feeds for children, will provide targeted protection to children to prevent encounters with illegal content.

14.16 These proposals are informed by our own assessment of the risks of harm, as set out in our Illegal Harms and Children's Register of Risks[1001] [1002], our Illegal Harms and Children's Risk Assessment Guidance and Risk Profiles,[1003] [1004] plus some additional sources as detailed in the chapters of this Statement. In each chapter we will explain how the measure will be effective in reducing risk and harm.

14.17 In relation to matters to which section 3(2)(g) in the 2003 Act is relevant, section 3(4A) sets out that in performing their duties under subsection (1), Ofcom must have regard to such of the following as appear to them to be relevant in the circumstances:

> (a) The risk of harm to citizens presented by content on regulated services.

14.18 As noted above, the Illegal Harms and Children's Register of Risks and Risk Profiles set out the risks of harm posed by content on regulated services. These risks, alongside findings from services' risk assessments, largely inform what measures will be appropriate for a service provider to address the risk of harm to citizens. The measures included in the Codes vary across services based on their risk and size. In each chapter we discuss the risk of harm that we are seeking to address and why we think our proposed measures will be effective.

[1000] Ofcom (2023), Impact assessment guidance, published 19 July 2023.
[1001] Ofcom (2026), Register of Risks, V2.0, published 25 June 2026.
[1002] Ofcom (2025), Children's Register of Risks, published 24 April 2025.
[1003] Ofcom (2026), Risk Assessment Guidance and Risk Profiles V2.0 Guidance published 25 June 2026.
[1004] Ofcom (2025), Children's Risk Assessment Guidance and Children's Risk Profiles published 24 April 2025.

(b) The need for a higher level of protection for children than for adults.

14.19 Our Codes already ensure a higher level of protection for children than for adults. In our amendments to the Codes, we are increasing the level of protection for children by introducing measures that will aid in targeting safety measures to protect children.

14.20 In this statement we explain our decision to recommend the use of highly effective age assurance to target safety default measures at children to address grooming and recommender system measures to protect children's recommender feeds. These measures will provide targeted protection against risks specific to children.

(c) The need for it to be clear to providers of regulated services how they may comply with their duties set out under the Act.

14.21 Our measures, and the explanation in this document about how our decisions and measures work, provide clarity and tangible steps that services can take to meet their duties in the Act. We recognise that in places we are not prescriptive. We have explained why that is in relation to the measures concerned and, where appropriate, prepared guidance to help providers abide by them. In each case, we are satisfied that the measure is clear enough.

14.22 In particular, in this statement, we are introducing draft Proactive technology Guidance to help providers with configuring and deploying proactive technology where applicable in the codes of practice. In this statement we also explain how we have updated existing guidance. We updated the Part 3 HEAA guidance[1005], to help providers understand how to use and implement age assurance where applicable to fulfil their duties under the Act. We also updated the Illegal Content Judgment Guidance to support service providers in judgments when they are unable to review content in messages, group chats, or forums. Finally, we updated our hash matching Guidance to give providers further direction on implementing hash matching measures.

(d) The need to exercise their functions so as to secure that providers of regulated services may comply with such duties by taking measures, or using measures, systems or processes, which are (where relevant) proportionate to (i) the size or capacity of the provider in question, and (ii) the level of risk of harm presented by the service in question, and the severity of the potential harm.

14.23 We have clearly identified in the Code measures the types and sizes of services we recommend adopt each measure, for the reasons given in each chapter of this statement.

14.24 The Act requires us to ensure measures are proportionate, and we recognise that the size, capacity, and risks of services differ widely. We therefore do not take a one-size-fits-all approach. Instead, we have set out what types of service we think should use specific safety measures to comply with their duties, with the most extensive expectations placed on the riskiest services.

14.25 Where appropriate we have recommended measures that give a degree of flexibility so that providers can tailor their approach to their individual service, taking into account factors including their size, capacity and risks on the individual service.

[1005] Ofcom (2025), Guidance on highly effective age assurance for Part 3 services.

(e) & (f) The desirability of promoting the use by providers of regulated services of technologies which are designed to reduce the risk of harm to citizens presented by content on regulated services; and the extent to which providers demonstrate, in a way that is transparent and accountable, that they are complying with their duties.

14.26 In this statement we are explaining our decision to recommend measures regarding the use of proactive technologies to detect suspected image-based CSAM content and terrorism content. The existing Codes (see ICU G1.2(b)) recommend the use of such technologies to be explained in providers' terms of service, as required by the Act.

14.27 Section 3(4) of the 2003 Act[1006] sets out other matters to which Ofcom must, to the extent they appear to us relevant in the circumstances, have regard, in performing our duties.

Section 3(4) : Ofcom must also have regard, in performing those duties, to such of the following as appear to them to be relevant in the circumstances […] (b) the desirability of promoting competition in relevant markets, (d) the desirability of encouraging investment and innovation in relevant markets; (h) the vulnerability of children and of others whose circumstances appear to Ofcom to put them in need of special protection; (i) the needs of persons with disabilities, of the elderly and of those on low incomes; (j) the desirability of preventing crime and disorder; (k) the opinions of consumers in relevant markets and of members of the public generally; (l) and the different interests of persons in the different parts of the United Kingdom, of the different ethnic communities within the United Kingdom and of persons living in rural and urban areas.

14.28 Where appropriate, in our measures, we have had regard to the desirability of promoting competition and encouraging investment and innovation. A number of our measures accordingly provide flexibility for service providers to decide how to achieve compliance. We recommend the use of classes of technology to detect image-based CSAM, incentivising providers of such technologies to offer cost-effective solutions. We are not prescriptive about which hash-matching databases providers should use, but only the criteria they must meet. As set out above, we have considered the interests of consumers in relevant markets as part of our impact assessments of proposed measures, including any indirect impacts on consumers in cases where our measures could affect competition, investment and innovation in respect of the online services that they use. To the extent that we are proposing measures related to the illegal content safety duties, these aim to prevent crime and disorder. We have had particular regard to the vulnerability of children in developing our recommendations on user banning for CSEA content and recommender systems. To properly consider the opinion of consumers in relevant markets we have taken into account feedback on our measures. In the chapters for the individual measures, we summarise the feedback received and how we have addressed it.

14.29 We consulted on our equality impact assessment and Welsh language assessment and have published our final assessments at annex 2: Equality impact assessment and Welsh language assessment. These assess the impact of the Codes on persons who may be vulnerable, including children and others whose circumstances appear to put them in need of special protection. Our equality impact assessment also has regard to the impact of the Codes on persons with disabilities, the elderly, those on low incomes and the different interests of persons in different parts of the UK, different ethnic communities within the UK and persons

[1006] As amended by section 82 of the Act.

living in rural and urban areas. This includes consideration of the impact of the Codes on opportunities to use the Welsh language and it being treated no less favourably than the English language.

14.30 In deciding on our measures to pursue the children's safety duties, we have had regard to the objective of a higher standard of protection for children than for adults, assessing whether measures are expected to be effective at achieving this.

Schedule 4: Online Safety Act 2003

14.31 As required by paragraph 1 of Schedule 4 to the Act, Ofcom has considered the appropriateness of provisions of the additional measures for the Codes of Practice to different kinds and sizes of Part 3 services and to providers of differing sizes and capacities and we have set out our reasons for our amendments to some Codes recommendations to services of different kinds, sizes and capacities.[1007]

14.32 We have had regard to the principles in Schedule 4 to the Act, as follows:

> Paragraph 2(a): providers of Part 3 services must be able to understand which provisions of the code of practice apply in relation to a particular service they provide[1008].

14.33 We have clearly identified which measures apply to what types and sizes, for the reasons given in each relevant chapter of this statement.

> Paragraph 2(b): the measures described in the code of practice must be sufficiently clear, and at a sufficiently detailed level, that providers understand what those measures entail in practice.[1009]

14.34 Having regard to the need for it to be clear to providers of regulated services how they may comply with their duties, we have been as clear as possible and included an appropriate level of detail in this statement and the supporting documents, consistent with acting proportionately. We have been sufficiently detailed and precise, while ensuring our measures are technically feasible and proportionate for the wide range of services in scope of the Act.

14.35 Our approach to the Codes strikes the balance between providing certainty about what providers need to do and allowing them flexibility to implement measures in a way that works in the context of their own services and is proportionate. For example, our hash matching measures give providers options on how to source hash matching databases. Additionally, the measures regarding user sanctions gives providers some flexibility to service providers on how to apply user sanctions.

14.36 Following stakeholder feedback and further proportionality assessments, we have taken steps to amend the wording of our user banning measures to ensure clarity in how providers should approach user banning for children in practice.

> Paragraph 2(c): the measures described in the code of practice must be proportionate and technically feasible: measures that are proportionate or technically feasible for providers of a certain size or capacity, or for services of a certain kind or size, may not be proportionate or

[1007] See also section 3(4A)(d) of the 2003 Act.
[1008] See also section 3(4A)(c) of the 2003 Act.
[1009] See also section 3(4A)(c) of the 2003 Act.

technically feasible for providers of a different size or capacity or for services of a different kind or size;

14.37 We have clearly identified in the Codes which measures apply to what types and sizes of services, for the reasons given in each chapter. We have considered proportionality and technical feasibility, where appropriate, as part of our impact assessments across this statement, including the combined impact assessment in chapter 13.

Paragraph 2(d): the measures described in the code of practice that apply in relation to Part 3 services of various kinds and sizes must be proportionate to Ofcom's assessment under section 98 of the risk of harm presented by services of that kind or size.

14.38 We have identified the relevant risks of harm that the measures address and explained why we consider each proposed measure is proportionate in the light of those harms. As required by section 3(4A)(b)(ii) of the 2003 Act, in considering proportionality we have had regard to the severity of the potential harm as well as the level of risk of harm, as identified in the Illegal Harms and Children's Register of Risks. Where appropriate, we have clearly identified which measures would apply to what types and sizes of services, for the reasons given in each relevant chapter of this statement.

14.39 Overall, the Codes place more demanding expectations on services that pose greater risks. Having regard to the desirability of encouraging investment and innovation in the markets for regulated services and these technologies, our recommendations provide sufficient flexibility for services. Our impact assessment for each measure, as well as our combined impact assessment, also take into account the cost to services as we acknowledge additional costs can affect investment and innovation. Additionally, we have determined in our combined impact assessment that the overall package of measures is proportionate to the cost for large services.

Ofcom's online safety objectives

User-to-User Services

14.40 As required by paragraph 3 of Schedule 4 to the Act, we have ensured that our recommendations are compatible with the pursuit of the applicable online safety objectives for user-to-user services as set out in this sub-section.

14.41 All our measures, in line with paragraph 11 of Schedule 4 to the Act, relate only to the design or operation of a Part 3 service (a) in the United Kingdom, or (b) as it affects United Kingdom users of the service.

Paragraph 4(a)(i): a service should be designed and operated in such a way that the systems and processes for regulatory compliance and risk management are effective and proportionate to the kind and size of service.

14.42 The existing Codes include measures related to governance and accountability, and in this statement, we have not changed these. Where appropriate in the interests of user safety, the measures we are now recommending include provisions on monitoring, record keeping and regular review.

Paragraph 4(a)(ii): a service should be designed and operated in such a way that the systems and processes are appropriate to deal with the number of users of the service and its user base.

14.43 In our December 2024 Statement at paragraph 14.14 and our April 2025 Statement at paragraph 21.53, we said that the content moderation, automated content moderation and reporting and complaints measure in our existing codes were set having regard, amongst other things, to the number of users of the service and its user base. To the extent we are recommending changes to these measures, or additional such measures, we consider these compatible with this objective.

14.44 We have recommended a number of measures which apply based in part on user-numbers, for the reasons set out in the relevant chapters.

Paragraph 4(a)(iii): a service should be designed and operated in such a way that United Kingdom users (including children) are made aware of, and can understand, the terms of service.

14.45 We are adding to our existing suite of measures relating to terms of service, measures recommending the inclusion in terms of service of information about sanctions against users. In our statement we explain how we have had regard to the importance of users understanding what sanctions are applicable if illegal content or content that is harmful to children is generated by a user of a service.

Paragraph 4(a)(iv): a service should be designed and operated in such a way that there are adequate systems and processes to support United Kingdom users.

14.46 As part of our measures, we are recommending additional reporting measures for livestreaming, to ensure it is easy for a UK user to report a livestream which features a risk of imminent physical harm. We are also proposing to amend the appeals mechanisms in the Codes to ensure they function effectively.

Paragraph 4(a)(vi): a service should be designed and operated in such a way that the service provides a higher standard of protection for children than for adults.

14.47 A number of existing measures are compatible with this objective. For example, user reporting and complaints measures ICU D2 and PCU D2 recommend that systems and processes should be easy to access and easy to find, helping child users report content harmful to children to the service provider.

14.48 Many of our additional measures and amendments in this statement are also compatible with this objective; for example, our measure recommending that service providers should use highly effective age assurance to target safety defaults to protect against grooming and protections relating to content recommender feeds at children.

Paragraph 4(a)(vii): a service should be designed and operated in such a way that the different needs of children at ages are taken into account.

14.49 Our additional measures recommend the use of highly effective age assurance on services, so that specific protections in relation to grooming and content recommendation can be targeted specifically at children.

14.50 We are also recommending that when implementing age assurance on a service, providers ensure that it is easy to use, including by children of different ages and with different needs.

Paragraph 4(a)(viii): a service should be designed and operated in such a way that there are adequate controls over access to the service by adults.

14.51 We have made amendments to the illegal content codes, to recommend that users who share CSEA material on a service are banned from the service, by having their access removed and being prevented from returning.[1010] Our measures also state that user-to-user services should have a policy regarding sanctions where UK users share other illegal content. We consider that the measures are consistent with this objective.

Paragraph 4(a)(ix): a service should be designed and operated in such a way that there are adequate controls over access to, and use of, the service by children, taking into account use of the service by, and impact on, children in different age groups.

14.52 We have updated the Illegal Content User-to-user Codes, with measures to recommend the use of highly effective age assurance on services for the purpose of complying with certain recommendations to protect children on the service.

Paragraph 4(b): a service should be designed and operated so as to protect individuals in the United Kingdom who are users of the service from harm, including with regard to:

- algorithms used by the service,
- functionalities of the service, and
- other features relating to the operation of the service.

14.53 Our measure in relation to recommender systems will ensure that content indicated as potentially priority illegal content is excluded from the recommender feeds of child users.

14.54 Our user control measures are being expanded to apply to more services and address more harms, to give more users access to functionalities and features, which can help them protect themselves from harms.

Search Services

14.55 We made limited changes to our search Codes. As required by paragraph 3 of Schedule 4 to the Act, we have ensured that our measures, in the context of our existing Codes, are compatible with the pursuit of the applicable online safety objectives as set out in this sub-section.

14.56 All the measures in the Codes, in line with paragraph 11 of Schedule 4 to the Act, relate only to the design or operation of a Part 3 service (a) in the United Kingdom, or (b) as it affects United Kingdom users of the service.

Paragraph 5(a)(i): a service should be designed and operated in such a way that the systems and processes for regulatory compliance and risk management are effective and proportionate to the kind and size of service.

14.57 In our December 2024 Statement and our April 2025 Statement, we have set out recommended measures that, amongst other things, have regard to the kind and size of a search service.

[1010] Consistent with paragraph 11 of Schedule 4 to the Act, this measure applies to the design and operation of the service in the UK, and as it affects UK users.

Paragraph 5(a)(ii): a service should be designed and operated in such a way that the systems and processes are appropriate to deal with the number of users of the service and its user base.

14.58 In our December 2024 Statement, April 2025 Statement and May 2026 Statement we have set out recommended measures that, amongst other things, are appropriate to deal with the number of users of the service and its user base.

Paragraph 5(a)(iii): a service should be designed and operated in such a way that United Kingdom users (including children) are made aware of, and can understand, the publicly available statement referred to in sections 27 and 29.

14.59 In our December 2024 Statement and our April 2025 Statement, we have set out recommended measures that, amongst other things, are designed to ensure the publicly available statement for a service is accessible to children.

Paragraph 5(a)(iv): a service should be designed and operated in such a way that there are adequate systems and processes to support United Kingdom users.

14.60 We are recommending measures to broaden content appeals relating to search content. We consider that this amendment to the Codes is in line with the objective.

Paragraph 5(a)(v): a service should be designed and operated in such a way that the service provides a higher standard of protection for children than for adults.

14.61 In our December 2024 Statement and our April 2025 Statement, we have set out recommended measures that support this goal. We note that the measures in the Protection of Children Search Code are designed to secure a higher level of protection for children than for adults.

Paragraph 5(a)(vi): a service should be designed and operated in such a way that the different needs of children at different ages are taken into account.

14.62 In our December 2024 Statement and our April 2025 Statement we set out recommended measures to support this goal including in the Reporting and Complaints and Publicly Available Statement chapters of our 2024 Statement and the draft Protection of Children Code for search services.

Paragraph 5(b): a service should be assessed to understand its use by, and impact on, children in different age groups.

14.63 As set out in our April 2025 Statement, service providers have a duty, as part of their children's risk assessment, to assess their user base, including the number of children in different age groups on the service. Additionally, service providers must assess the impact of the risk of harm to children in different age groups on their services.

Paragraph 5(c): a search engine should be designed and operated so as to protect individuals in the United Kingdom who are users of the service from harm, including with regard to:

- algorithms used by the search engine,
- functionalities relating to searches (such as a predictive search functionality), and
- the indexing, organisation and presentation of search results

14.64 We consider our recommendation in our May 2026 Statement to be compatible with this objective, alongside our recommendations in Volume 1, Chapter 5; Volume 2, Chapter 3; and Volume 2, Chapter 9 of our December 2024 Statement, and Section 11, Section 15 and Section 19 of our April 2025 Statement.

Schedule 4 requirements on Content of Codes of practice

User-to-User Services

14.65 Codes of Practice that describe measures recommended for the purpose of compliance with a duty set out in section 10(2), 10(3), or 10(3A) (illegal content) and section 12(2) or 12(3) (safety duties protecting children) of the Act, must include measures in each of the areas of a service listed in section 10(4) (illegal content) and section 12(8) (safety duties protecting children). This provision applies to the extent that inclusion of the measures in question is consistent with:

- Ofcom's duty to consider the appropriateness of provisions of the Code of Practice to different kinds and sizes of Part 3 services and to providers of differing sizes and capacities;
- the principle that the measures described in the Code of Practice must be proportionate and technically feasible; and
- the principle that the measures described in the Code of Practice that apply in relation to Part 3 services of various kinds and sizes must be proportionate to Ofcom's assessment (under section 98) of the risk of harm presented by services of that kind or size.

14.66 In our December 2024 statement (volume 2, paragraph 14.18) and April 2025 statement (volume 4, paragraph 21.94) we explained that our user-to-user Codes included recommendations in each of the areas of a service listed in sections 27(4) for search services in the following areas of a service listed in section 27(4) and 29(4). To the extent we are making further recommendations they are in the following areas:

- Design of functionalities, algorithms and other features: measures in relation to recommender systems.
- Policies on user access to the service or to particular content present on the service, including blocking users from accessing the service or particular content: measures relating to the removal of access for users who share CSEA material, and a wider user sanctions measure.
- Content moderation, including the use of automated tools to detect content for moderation and measures to secure that action taken in error against content is reversed to the extent possible and appropriate, even if the user does not appeal.
- User support measures: improved reporting functions for livestreaming, a broader appeals measure, and the use of HEAA to target existing measures to address grooming in the Illegal Content Codes of Practice.

14.67 The measures in this statement have been assessed (see individual chapters) for their impact on users' rights in line with Paragraph 10(1)-(3) of Schedule 4 to the Act which requires

measures described in a code of practice, which are recommended for the purpose of compliance with any of the relevant duties, to be designed in the light of the following principles:

- The importance of protecting the rights of users and (in the case of combined services) interested persons to freedom of expression within the law.
- The importance of protecting the privacy of users.

Search Services

14.68 Codes of Practice that describe measures recommended for the purpose of compliance with a duty set out in section 27(2), 27(3), or 27(3)(a) (illegal content) and section 29(2) or 29(3) (safety duties protecting children) of the Act, must include measures in each of the areas of a service listed in section 27(4) (illegal content) and section 29(4) (safety duties protecting children).

14.69 This provision applies to the extent that inclusion of the measures in question is consistent with:

- Ofcom's duty to consider the appropriateness of provisions of the Code of Practice to different kinds and sizes of Part 3 services and to providers of differing sizes and capacities;
- the principle that the measures described in the Code of Practice must be proportionate and technically feasible; and
- the principle that the measures described in the Code of Practice that apply in relation to Part 3 services of various kinds and sizes must be proportionate to Ofcom's assessment (under section 98) of the risk of harm presented by services of that kind or size.

14.70 In our December 2024 statement (volume 2, paragraph 14.23) and April 2025 statement (volume 4, paragraph 21.98) we explained that our Search Codes included recommendations in each of the areas of a service listed in sections 27(4) for search services in the following areas of a service listed in section 27(4) and 29(4). To the extent we are making further recommendations in relation to the design of functionalities and features relating to search engines, see broadening content appeals chapter. We have also recommended a measure to secure that action taken in error against content is reversed to the extent possible and appropriate, even if the interested person does not appeal.

Schedule 4 requirements on Content of Codes of Practice: age assurance

14.71 Our measures recommend the use of age assurance to help providers accurately determine which UK users are child users and apply appropriate safeguards to protect against serious illegal harms impacting children, including grooming, child sexual exploitation and abuse, content encouraging or assisting suicide and serious self-harm, and terrorism and/or hate.

14.72 Paragraph 12 of Schedule 4 to the Act, sets out the requirements of the age assurance measures of the Codes of Practice. Paragraph 12(1) of Schedule 4 to the Act describes the inclusion of age assurance in a code of practice as a measure recommended for the purpose of compliance with any of the duties set out in section 12(2) or (3) or 29(2) or (3) of the Act.

14.73 Though our measures are being recommended for the purpose of compliance with section 10(2) and 10(3) of the Act, the focus of these measures is on protecting children from severe

risks of illegal harm. Therefore, we have taken steps in line with paragraph 12 of Schedule 4 to the Act.

> Paragraph 12(2): In deciding whether to recommend the use of age assurance, or which kinds of age assurance to recommend, Ofcom must have regard to the following:
>
> a) the principle that age assurance should be effective at correctly identifying the age or age-range of users
>
> b) relevant standards set out in the latest version of the code of practice under section 123 of the Data Protection Act 2018 (age-appropriate design code);
>
> c) the need to strike the right balance between: (i) the level of risk and the nature, and severity, of potential harm to children which the age assurance is designed to guard against, and (ii) protecting the right of users and (in the case of search services or the search engine of combined services) interested persons to freedom of expression within the law; ages and with different needs;
>
> d) the principle that more effective kinds of age assurance should be used to deal with higher levels of risk of harm to children;
>
> e) the principle that age assurance should be easy to use, including by children of different
>
> f) the principle that age assurance should work effectively for all users regardless of their characteristics or whether they are members of a certain group;
>
> g) the principle of interoperability between different kinds of age assurance.

14.74 In chapter 5: Highly effective age assurance in the Illegal Content Codes

14.75 and the chapters[1011] where we are recommending the use of highly effective age assurance to target specific measures, we discuss our measures for user-to-user services regarding age assurance, including how we have had regard to the factors set out in paragraph 12(2)(a) – (2)(g) of Schedule 4 (see above) in developing the policy that has informed the measures. We discuss the severity of the harm that age assurance is being used to protect children against, and how we have balanced that against users' freedom of expression and privacy rights.

14.76 We have adopted measures that are supported by the draft Guidance for Part 3 services on highly effective age assurance (Part 3 HEAA Guidance), to ensure that age assurance is easy to use, is effective regardless of the user who uses it, and to promote the principle of interoperability between age assurance methods.

14.77 The Part 3 HEAA Guidance supports all service providers in understanding whether any age assurance they have in place is highly effective and supports user-to-user services in understanding how they may implement highly effective age assurance for the purpose of implementing the recommended Age Assurance measures.

[1011] Chapter 6: Increasing the effectiveness of safety defaults and supportive messaging and chapter 11: Recommender systems.