Protecting people from illegal harms online

Annex 1: Further stakeholder responses to our November 2023 Illegal Harms Consultation

Statement

Published 16 December 2024

Contents

A1. Further stakeholder responses

About this document

A1.1.1 This annex addresses the additional points that were made by respondents to our November 2023 Illegal Harms Consultation, our August 2024 Further Consultation on Torture and Animal Cruelty, and our May 2024 Consultation on Protecting children from harms online.

A1.1.2 While we have reviewed all the relevant feedback, we did not consider that the points set out in this Annex required substantial additions or explanations in our Statement.

A1.2 Our approach to developing Codes measures

Stakeholder responses by theme

Safety by design

A1.2.1 Just Algorithms Action Group said there were several different uses of the term 'safer' in our November 2023 Illegal Harms Consultation ('November 2023 Consultation').[1]

Our response

A1.2.2 The intention of the Online Safety Act 2023 ('the Act') is to protect users online and ensure the many versions of safety for these users – including safety from encountering harmful content and psychological safety. There is therefore a need to refer to many different versions of the term.

Services

A1.2.3 [✂] argued that we should provide exceptions to the application of Codes measures were user-to-user (U2U) content represents an ancillary function for a service and/or is not the primary offering of the service.[2]

Our response

A1.2.4 We would not look to provide an exception for this reason. We outlined in our Overview of regulated services chapter where they may be exemptions under the Act.

Timing and timelines

A1.2.5 eBay asked we recognise the major constraints associated with product changes, noting that our current timelines allow little time for services to adapt in the event that our final guidance and measures differ significantly from those presented during the November 2023 Consultation.[3] Similar points about the need for clarity on timeline were made in response to our May 2024 Consultation on Protecting Children from Harms Online ('May 2024 Consultation').[4]

Our response

A1.2.6 We address this point in Volume 3: chapter 3: Ofcom's enforcement powers.

---

[1] Just Algorithms Action Group response to November 2023 Illegal Harms Consultation pp.1-2.
[2] [✂]
[3] eBay response to November 2023 Illegal Harms Consultation, pp.3-4.
[4] Federation of Small Businesses response to May 2024 Consultation on Protecting Children from Harms Online, pp.1-2; Meta response to May 2024 Consultation on Protecting Children from Harms Online, p.16; Pinterest response to May 2024 Consultation on Protecting Children from Harms Online, pp.5-6; Yoti response to May 2024 Consultation on Protecting Children from Harms Online, p.24.

Costs of measures

A1.2.7 The Mid Size Platform Group raised concerns that the prices of technologies used for our recommended measures may become inflated, which could make implementing them financially unfeasible for providers of smaller services. Additionally, it asked for a cumulative impact assessment in cases where providers need to apply several measures.[5]

Our response

A1.2.8 At present, we do not have enough evidence to show that the costs of implementing specific measures would rise significantly due to the limited capacity of technology suppliers. Costs could also decrease due to increased competition from new suppliers of specific technologies. More broadly, we expect any impact from limited technology supplier capacity is likely to be temporary, because we would expect increased capacity to be developed in response to increasing demand.

A1.2.9 We have carried out a combined impact assessment, which is available at Volume 2: Chapter 13. The cumulative costs to a provider will depend on the particular measures that apply to it. This will vary with functionalities of its service, the risk levels for different kinds of illegal harm and the service size. We have not set out all permutations of this.

Costs of compliance

A1.2.10 Some stakeholders raised concerns over the costs of compliance with the Act, highlighting that these were likely to be significant.[6]

Our response

A1.2.11 We recognise that there is a baseline level of compliance costs that exists for all service providers. As described in Volume 2: Chapter 13, for providers of smaller services that assess as low risk for all kinds of harm, most of the measures we impose relate to specific requirements in the Act over which we have no discretion. We impose very few measures for such services beyond those specifically required by the Act. For services with higher risks, the measures are more demanding and costly to implement, but we assess them to be proportionate. When considering whether to impose measures, we have considered the costs to services that might result from their implementation.

Benefits

A1.2.12 In response to the November 2023 Consultation, some stakeholders argued that we underestimated the benefits from applying a measure to reduce harms online because we did not consider wider impacts on society (such as those on the justice system or on the public health system).[7] The Molly Rose Foundation questioned the use of the "standard UK

[5] MSPG response to November 2023 Illegal Harms Consultation, p.7.
[6] Online Dating and Discovery Association response to November 2023 Illegal Harms Consultation, p.2; SafeCast response to November 2023 Illegal Harms Consultation, p.7.
[7] International Justice Mission response to November 2023 Illegal Harms Consultation, p.13; Molly Rose Foundation response to November 2023 Illegal Harms Consultation, pp.8-9.

model for assessing value of a prevented fatality" and suggested that we use the J-value method.[8] It argues for using the J-value method to assess the benefits of reducing suicide and self-harm.

Our response

A1.2.13 For each of our measures, we have qualitatively assessed the benefits arising from preventing harms online (including those related to suicide and self-harm) since there are wider societal impacts of harms or longer-term impacts which are difficult to reliably quantify in monetary terms. For some measures, we have discussed the benefits of reducing CSEA harms in terms of monetary value (see Annex 5). While this monetary estimate includes several social costs (such as the costs associated with policing), we have acknowledged that it may not be possible to assign a monetary value to all possible benefits that flow from reducing CSEA harms. As such, we have combined our assessment of monetary benefits with qualitative reasoning.

Disagreement with the Act

A1.2.14 Big Brother Watch criticised the Act, calling it "flawed".[9] BitChute also expressed objections to the Act and its approach to illegal content.[10] One respondent said "The UK Online Safety Bill was also compiled without the involvement of core groups – sex workers, pornography performers and pornography producers."[11]

Our response

A1.2.15 It is not within Ofcom's powers to amend the Act.

[8] The Value of Prevented Fatality (VPF) is used by the Department for Transport, which is referred to in the HM Treasury's Green Book. The current UK method for VPF gives a single value for all ages, while the proposed J-value method accounts for the difference in benefit to people of varying age by considering variations in the life expectancy.

[9] Big Brother Watch response to November 2023 Illegal Harms Consultation, p.1.

[10] Bitchute response to November 2023 Illegal Harms Consultation, p.1.

[11] Name Withheld 1 response to November 2023 Illegal Harms Consultation, p.4.

A1.3 Register of Risks

Introduction

A1.3.1 In this Annex, we set out and explain the decisions we have made in finalising the Illegal Harms Register of Risks ('Register of Risks'). In particular, we explain how we have taken into account stakeholder feedback received in response to our November 2023 Illegal Harms Consultation, our August 2024 Further Consultation on Torture and Animal Cruelty, and our May 2024 Consultation on Protecting children from harms online.

A1.3.2 We have structured this Annex by including one section per chapter of the Register of Risks.

General responses

Summary of stakeholder responses

A1.3.3 We received responses from stakeholders on the overall approach of the Register of Risks. Stakeholders who provided information in their consultation responses that was reviewed in this area included service providers (Booking.com[12], [✂][13], DuckDuckGo[14], Apple[15], Roblox[16] , [✂][17], [✂][18], [✂][19], LinkedIn[20], Proton[21], Nexus[22]), academics (SPRITE+[23], Oxford Disinformation and Extremism[24]), non-profit organisations (Marie Collins Foundation[25], Centre for Competition Policy[26], CELE[27], Mencap[28], UK Interactive Entertainment[29], techUK[30], Electronic Frontier Foundation[31], Global Network Initiative[32],

[12] Booking.com response to the Illegal Harm November 2023 Consultation, pp. 3-11.
[13] [✂]
[14] DuckDuckGo response to the Illegal Harm November 2023 Consultation, p. 2.
[15] Apple response to the Illegal Harm November 2023 Consultation, pp. 2-3.
[16] Roblox response to the Illegal Harm November 2023 Consultation, p. 2.
[17] [✂]
[18] [✂]
[19] Name withheld 5 response to the Illegal Harm November 2023 Consultation, pp. 1-2.
[20] LinkedIn response to the Illegal Harm November 2023 Consultation, pp. 1-2.
[21] Proton response to the Illegal Harm November 2023 Consultation, pp. 2-3.
[22] Nexus response to the Illegal Harm November 2023 Consultation, pp. 1-2.
[23] SPRITE+ response to the Illegal Harm November 2023 Consultation, p. 2.
[24] Oxford Disinformation and Extremism response to the Illegal Harm November 2023 Consultation, p. 2.
[25] Marie Collins Foundation response to the Illegal Harm November 2023 Consultation, pp. 2-3.
[26] Centre for Competition Policy response to the Illegal Harm November 2023 Consultation, pp. 2-5.
[27] CELE response to the Illegal Harm November 2023 Consultation, pp. 3-4.
[28] Mencap response to the Illegal Harm November 2023 Consultation, p. 2.
[29] UK Interactive Entertainment response to the Illegal Harm November 2023 Consultation, p. 2.
[30] techUK response to the Illegal Harm November 2023 Consultation, pp. 6-8.
[31] Electronic Frontier Foundation response to the Illegal Harm November 2023 Consultation, pp. 1-2.
[32] Global Network Initiative response to the Illegal Harm November 2023 Consultation, p. 2.

Just Algorithms Action Group[33], Canadian Center for Child Protection[34], Global Partners Digital[35], CARE[36], Global Encryption Coalition[37], South East Fermanagh Foundation[38], Trust Alliance Group[39]); civil society organisations (OSAN[40], NWG Network[41]); public bodies (Scottish Government[42]); individuals (Name Withheld 3[43]); private companies (Segregated Payments Ltd[44], Innovate Finance[45], INVIVIA Inc.[46]).

A1.3.4 We have decided to make following changes of each type to the relevant Register of Risks chapter.

Changes or additions we have made within the Register of Risks

A1.3.5 We have adjusted our definition of content recommender systems to clarify that content recommender systems which only recommend to a user their own user-generated content are not in scope of this definition. See Register of Risks glossary for the updated definition.

A1.3.6 We have clarified that all references to 'content recommender systems' throughout the Register of Risks do not encompass 'product recommender systems'. Under the Online Safety Act 2023 ('the Act'), the definition of user-generated content is inclusive of user-generated product listings on online U2U marketplaces. However, we have carved out recommender systems used exclusively for the purposes of recommending goods and services for sale and hire on online marketplaces and listing services from our definition of content recommender systems. We refer to these as 'product recommender systems'. Product recommenders are technically distinct from content recommender systems, but the two technologies are not legally distinguishable due to the definition of content in the Act being inclusive of product listings. There are two reasons for carving out product recommender systems from our definition of content recommender systems. First, evidence of product recommender systems contributing to the dissemination of illegal content is lacking. Second, they are a distinct technology which would require a distinct policy approach independent of content recommender systems.

A1.3.7 We have expanded on the benefits of end-to-end encryption (E2EE) in the Introduction of the Register of Risks to ensure this presents a balanced view of the functionality. We have

[33] Just Algorithms Action Group response to the Illegal Harm November 2023 Consultation, p. 5.
[34] Canadian Center for Child Protection response to the Illegal Harm November 2023 Consultation, pp. 2-4.
[35] Global Partners Digital response to the Illegal Harm November 2023 Consultation, pp. 5-8.
[36] CARE response to the Illegal Harm November 2023 Consultation, pp. 2-5.
[37] Global Encryption Coalition response to the Illegal Harm November 2023 Consultation, pp. 1-2.
[38] South East Fermanagh Foundation response to the Illegal Harm November 2023 Consultation, pp. 2-3.
[39] Trust Alliance Group response to the Illegal Harm November 2023 Consultation, pp. 1-2.
[40] OSAN response to the Illegal Harm November 2023 Consultation, pp. 8-100.
[41] NWG Network response to the Illegal Harm November 2023 Consultation, pp. 3-7.
[42] Scottish Government response to the Illegal Harm November 2023 Consultation, pp. 2-4.
[43] Name Withheld 3 response to the Illegal Harm November 2023 Consultation, pp. 1-2.
[44] Segregated Payments Ltd response to the Illegal Harm November 2023 Consultation, pp. 2-3.
[45] Innovate Finance response to the Illegal Harm November 2023 Consultation, p. 3.
[46] INVIVIA Inc response to the Illegal Harm November 2023 Consultation, pp. 2-3.

further clarified in the Introduction of the Register of Risks that we do not consider any characteristics of a service inherently harmful, but that the role of the Register of Risks is to highlight where characteristics appear to heighten a risk of harm.

A1.3.8 We have added further evidence on Generative AI (GenAI) to our Register of Risk, as more evidence on GenAI has been produced since our consultation and various stakeholders raised the importance of acknowledging how it can play a role in facilitating specific kinds of illegal harm as well as more general risks associated with the advancement and accessibility of GenAI.

Introduction: Business model and commercial profiles

Summary of stakeholder responses

A1.3.9 Stakeholders who provided information in their consultation response relevant to this area included service providers, and civil society and other independent organisations (5Rights Foundation[47], Centre for Competition Policy[48], Wikimedia Foundation[49], Molly Rose Foundation[50], Proton[51], Sprite+[52], TrustAllianceGroup[53], (DuckDuckGo)[54], Global Partners Digital[55], Spotify[56], and Global Disinformation Index[57]).

A1.3.10 Responses relevant to the Register of Risks focused on a wide variety of issues and evidence. We received a number of suggestions for amendments. This feedback has been used to update our text on the evidence base linking Business Model and Commercial Profiles to harms.

Further clarity provided to our conclusions on how evidence is articulated

A1.3.11 Although we highlight where there is evidence that links specific harms with business model characteristics in relevant harms chapters, we have included a broader discussion of business models and commercial profiles within the Introduction of the Register of Risks. Additional evidence and recommendations from stakeholders regarding the need for further emphasis on the role they play in increasing the risk of various harms has largely been addressed in this introductory section.

[47] 5Rights Foundation response to the Illegal Harm November 2023 Consultation, pp. 4-7.
[48] Centre for Competition Policy response to the Illegal Harm November 2023 Consultation, p. 3.
[49] Wikimedia Foundation response to the Illegal Harm November 2023 Consultation, pp. 1-6.
[50] Molly Rose Foundation response to November 2023 Illegal Harms Consultation, pp. 26-28.
[51] Proton response to November 2023 Illegal Harms Consultation, p. 2.
[52] Sprite+ response to November 2023 Illegal Harms Consultation, p. 5.
[53] TrustAllianceGroup response to November 2023 Illegal Harms Consultation, p. 2.
[54] DuckDuckGo response to the Illegal Harm November 2023 Consultation, p. 3.
[55] Global Partners Digital response to November 2023 Illegal Harms Consultation, pp. 7-8.
[56] Spotify response to November 2023 Illegal Harms Consultation, pp. 3-5.
[57] Global Disinformation Index response to November 2023 Illegal Harms Consultation, pp. 2-3.

A1.3.12 We have provided further commentary in regard to how the incentive to optimise revenue may lead a service provider to implement or favour functionalities or features to maximise user engagement (e.g. through recommender systems) at the expense of other outcomes, such as user safety (across harms).

A1.3.13 We have expanded the introductory section to provide clarity that commercial incentives may not sufficiently support the development of systems to minimise the risk of harms to individuals as this may lead to revenue or profit reduction.

A1.3.14 We have expanded our discussion in regard to revenue models, explaining how various revenue models (not necessarily limited to advertising-based or subscription-based models) may result in financial incentives that cause providers to prioritise user engagement over user safety.

Terrorism

Summary of stakeholder responses

A1.3.15 Responses relevant to this chapter of the Register of Risks focused predominantly on highlighting new evidence which related to risk factors not currently set out in the Register of Risks, including those related to user base and specific platform functionalities. Some responses also focused on additional research to bolster our existing evidence base.

A1.3.16 Stakeholders who provided or cited new research, evidence, or recommendations in this area included academics (Dr. Sandy Schumann, University College London[58]); and independent organisations and individuals (Children's Commissioner[59], Institute for Strategic Dialogue[60], Glitch[61], Jonathan Hall (Independent Reviewer of Terrorism Legislation)[62], 5Rights Foundation[63] and Tech Against Terrorism[64]).

A1.3.17 We have decided to make following changes of each type to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.18 We have added the risk factor, 'age', specifically in relation to children being at increased risk of radicalisation due in part to the amount of time they spend online, which is likely to have been exacerbated as a result of social isolation during the COVID-19 pandemic. We

[58] Dr. Sandy Schumann response to November 2023 Illegal Harms Consultation, pp. 3-5.
[59] Children's Commissioner response to November 2023 Illegal Harms Consultation, p. 20
[60] Institute for Strategic Dialogue response to November 2023 Illegal Harms Consultation, pp. 4-5.
[61] Glitch response to November 2023 Illegal Harms Consultation, pp. 2-3.
[62] Jonathan Hall response to the Illegal Harm November 2023 Consultation, pp. 1-4.
[63] 5Rights Foundation response to November 2023 Illegal Harms Consultation, p. 7.
[64] Tech Against Terrorism response to November 2023 Illegal Harms Consultation, pp. 2-6.

have included research from the European Union Situation and Trend Report 2022 and Home Office statistics relating to the Prevent programme.

A1.3.19 We have added the risk factor, 'user-generated content searching', in relation to the increased risk of encountering terrorism content online where users actively seek out such content.

A1.3.20 We have added further evidence for the risk factor, 'posting content (text, images, videos)', in reference to the use of original audio content being added to videos as a way for proscribed groups to spread propaganda, as identified by The Institute for Strategic Dialogue.

A1.3.21 We have added further evidence to the risk factor, 'file-storage and file-sharing services', with corresponding evidence provided by stakeholders in their consultation responses for risks associated with 3D printed firearms offences in particular.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.22 Based on Home Office advice, we have aligned some of our definitions of terrorism and terrorist activities with legislation, such as the Terrorism Act 2000.

A1.3.23 We have added further commentary regarding the role of generative AI in creating terrorism content, using evidence from Tech Against Terrorism.

A1.3.24 We have provided further clarity around the figures used to demonstrate the volume of UK users assumed to be exposed to content encouraging extremism, terrorism or radicalisation. We have more clearly explained the limits of the definitions used in research and included additional statistics using different definitions for comparison.

A1.3.25 We have provided further clarity on the role of race and other personal characteristics in relation to the likelihood of users encountering, or being targeted by, online terrorism content, citing additional evidence and new Ofcom data.

A1.3.26 We have added a clarification to evidence that shows 3D printing has been used to enable the conversion of blank-firing firearms into live-firing firearms.

Proposed changes we are not making despite requests to do so

A1.3.27 We received feedback from Glitch stating there should be there should be more effort towards exploring the intersection between terrorism and the grooming of young girls. However, since there was no corresponding evidence provided and there appears to be a lack of evidence on this topic, we have decided to not update the chapter.

Child Sexual Exploitation and Abuse (CSEA); including Grooming and Child Sexual Abuse Material (CSAM)

Summary of stakeholder feedback

A1.3.28 Stakeholders provided feedback and a range of new sources which have been used to update the CSEA chapters, including specific sub-chapters on grooming and CSAM.

A1.3.29 Stakeholders who provided feedback and new evidence as part of their consultation responses relevant to this chapter included civil society and other independent organisations or individuals (Age Verification Providers Association[65], Barnardo's[66], Canadian Center for Child Protection (C3P)[67], CARE[68], Centre of Expertise on Child Sexual Abuse,[69] Cyacomb Ltd.[70], Cybersafe Scotland[71], Electronic Frontier Foundation[72], GeoComply Solutions[73], Glitch[74], Lucy Faithfull Foundation[75], The Independent Inquiry into Child Sexual Abuse (IICSA)[76], International Justice Mission's Center to End Online Sexual Exploitation of Children[77], Internet Matters[78], Internet Watch Foundation (IWF)[79], Marie Collins Foundation[80], NSPCC[81], Online Safety Act Network[82], Philippines Survivor Network[83], Protection Group International, SafeCast[84], UK Safer Internet Centre (UKSIC)[85],

[65] Age Verification Providers Association response to November 2023 Illegal Harms Consultation. p. 2.
[66] Barnardo's response to November 2023 Illegal Harms Consultation. pp. 4-6.
[67] Canadian Center for Child Protection (C3P) response to November 2023 Illegal Harms Consultation. pp. 2-3.
[68] CARE response to November 2023 Illegal Harms Consultation. pp. 3-4.
[69] Centre of Expertise on Child Sexual Abuse response to November 2023 Illegal Harms Consultation. pp. 2-8.
[70] Cyacomb Ltd. response to November 2023 Illegal Harms Consultation. pp. 3-4.
[71] Cybersafe Scotland response to November 2023 Illegal Harms Consultation. p. 1.
[72] Electronic Frontier Foundation response to November 2023 Illegal Harms Consultation. p. 1.
[73] GeoComply Solutions response to November 2023 Illegal Harms Consultation. pp. 1-4.
[74] Glitch response to November 2023 Illegal Harms Consultation. pp. 2-5.
[75] Lucy Faithfull Foundation response to May 2024 Protection of Children Consultation, p. 7.
[76] The Independent Inquiry into Child Sexual Abuse (IICSA) response to November 2023 Illegal Harms Consultation. pp. 1-2.
[77] International Justice Mission's Center to End Online Sexual Exploitation of Children response to November 2023 Illegal Harms Consultation. pp. 1-3, 5.
[78] Internet Matters response to November 2023 Illegal Harms Consultation. pp. 3-9.
[79] Internet Watch Foundation response to November 2023 Illegal Harms Consultation. pp. 3-9, 25, 30-31.
[80] Marie Collins Foundation response to November 2023 Illegal Harms Consultation, pp. 2-3.
[81] NSPCC response to November 2023 Illegal Harms Consultation, pp. 1-5.
[82] Online Safety Act Network response to November 2023 Illegal Harms Consultation, pp. 80, 99.
[83] Philippines Survivors Network response to November 2023 Illegal Harms Consultation, p. 2.
[84] Safecast response to November 2023 Illegal Harms Consultation. pp. 2.
[85] UK Safer Internet Centre (UKSIC) response to November 2023 Illegal Harms Consultation, pp. 23-25, 42.

WeProtect[86], Zevo Health[87], Scottish Government[88], [✂ [89]); service providers [✂][90], [✂][91], Mid Size Platform Group[92]).

A1.3.30 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

Grooming

A1.3.31 We have added further evidence to the risk factor, 'messaging services', drawing on evidence provided by stakeholders, particularly that which outlines the growth of online grooming during the COVID-19 pandemic.

A1.3.32 We have added further commentary and evidence to the user base risk factors sections on 'disability', 'sexual orientation and gender identity', 'socio-economic factors' and 'ethnicity', drawing on evidence and recommendations provided by stakeholders to further address personal circumstances which can lead to children being more vulnerable or be at 'higher risk'. This highlights, for example, the increased risk children with special educational needs and disabilities face in relation to grooming.

A1.3.33 We have added the risk factor, 'anonymous user profiles', drawing on more recent research that highlighted the importance of this risk factor in the experiences of children and young people who have received sexualised messages from users perceived to be adults.

A1.3.34 We have added further evidence to the sub-section, 'user connections', specifically the user networking risk factor, drawing on evidence provided by stakeholders and new research.

A1.3.35 Further evidence has been added to the 'fake user profiles' and 'anonymous user profiles' risk factors, which demonstrates that anonymised user profiles are used to send explicit images to children.

A1.3.36 We have expanded the sub-section on 'user connections' to further explain how perpetrators approach their victims and how the nature of messages can evolve in grooming dynamics.

A1.3.37 We have added further evidence to the risk factors, 'direct messaging' and 'encrypted messaging', which outlines the use of these functionalities to send sexualised messages to children and young people, drawing on evidence provided by stakeholders.

---

[86] WeProtect Global Alliance response to November 2023 Illegal Harms Consultation, pp. 2-5.
[87] Zevo Health response to November 2023 Illegal Harms Consultation, p. 2.
[88] Scottish Government response to November 2023 Illegal Harms Consultation, pp. 2-3.
[89] [✂]
[90] [✂]
[91] [✂]
[92] Mid-Size Platform Group response to November 2023 Illegal Harms Consultation, p. 3.

A1.3.38 We have added the 'ephemeral messaging' risk factor in response to evidence provided by stakeholders that highlights how this specific functionality can play a particular role in way that perpetrators interact, or attempt to interact, with children on certain services.

CSAM

A1.3.39 We have added further evidence and commentary to the risk factors, 'social media services' and 'video-sharing services', to include additional evidence provided by stakeholders that conveys the rate at which users encounter, or actively view, search for, or share CSAM online.

Further clarity provided to our conclusions or how evidence is articulated

CSEA (overview)

A1.3.40 We have added further information on the demographics of perpetrators of CSEA offences, in response to feedback that this is important information to understand how the offences manifest.

A1.3.41 We have added a new section, 'cross-cutting harm', to explain the overlapping nature of CSEA and other harms (such as intimate image abuse and suicide). Within this section we have included further evidence and commentary covering the topics of self-generated indecent imagery (SGII), financially motivated sexual extortion ('sextortion'), other egregious harms and the role of technology such as GenAI in further enabling and facilitating both grooming and CSAM offences.

CSAM

A1.3.42 We have updated various statistics related to the scale of CSAM and CSAM offences in the UK, drawing on new evidence provided by stakeholders and updated research and reporting identified since the Consultation.

A1.3.43 We have provided more granular commentary and evidence regarding the user base demographics and the characteristics of users that might increase the risks of them being victims of CSAM offences, in response to evidence provided by stakeholders and requests to ensure we were capturing these distinctions accurately.

Proposed changes we are not making despite requests to do so

A1.3.44 We have decided not to further categorise the harms associated with CSAM at a more granular level, as we believe our current separation of the harm encompasses all the types of harms in the area.

Hate

Summary of stakeholder feedback

A1.3.45 Responses relevant to this chapter of the Register of Risks focused on a wide variety of issues and new evidence. Stakeholders provided insights into the range of relevant protected characteristics, as well as how these characteristics interrelate. The chapter has been updated to reflect the range of feedback and evidence we have received. Additionally, this chapter has been expanded to encompass new evidence of online hate that has resulted from the Southport Riots.

A1.3.46 Stakeholders who provided information in their consultation responses relevant to this area include civil society (Humanists UK[93], The Board of Deputies of British Jews[94], South East Fermanagh Foundation[95], Coventry Youth Activists[96]); and other independent organisations or individuals (Institute for Strategic Dialogue[97], Glitch[98]).

A1.3.47 The following changes of each type have been made to the Register of Risks chapter 'Hate' for publication.

Changes or additions to the evidence base referencing risk factors

A1.3.48 We have added to our evidence base to the risk factor, 'social media services', to highlight how social media services can be used to promote hateful ideologies and direct targeted hate towards communities.

A1.3.49 We have added the risk factor, 'video-sharing services', with corresponding evidence that shows the prevalence of hateful content on these services and how they are used in the context of specific events to proliferate such content. This change was in response to Institute of Strategic Dialogue's feedback to our November 2023 Illegal Harms Consultation.

A1.3.50 We have added further evidence to the risk factor, 'user base demographics', on how religion, disability, sexuality, and the intersectionality of such personal characteristics are risk factors for hate.

A1.3.51 We have added further evidence to the risk factor, 'user-identification', to show the role of pseudonymous and anonymous user profiles in committing hateful offences.

[93] Humanists UK response to November 2023 Illegal Harms Consultation, pp. 1-6.
[94] The Board of Deputies of British Jews response to November 2023 Illegal Harms Consultation, p. 2.
[95] South East Fermanagh Foundation response to November 2023 Illegal Harms Consultation, pp. 2-3.
[96] Ofcom / Coventry Youth Activists meeting, 16 April 2024.
[97] Institute for Strategic Dialogue response to November 2023 Illegal Harms Consultation, pp. 3-5.
[98] Glitch response to November 2023 Illegal Harms Consultation, p. 3.

A1.3.52 We have added further evidence to the risk factor, 'user-communications', to show how comments can be used to spread hateful content online, especially during real-world events.

A1.3.53 We have added further evidence to the risk factor, 'reacting to content', which demonstrates how individuals are encouraged to post hateful content due to the status it brings them, particularly within communities of like-minded individuals.

A1.3.54 We have added further evidence to the risk factor, 'posting content', to highlight how services that enable users to post content can be used to spread hate, especially during real-world events.

A1.3.55 We have added the risk factor, 'user groups', with corresponding evidence to support that services that allow users to build online communities can enable hate offenders to spread hateful content amongst like-minded users.

A1.3.56 We have included further commentary to the risk factor, 'advertising-based revenue model', to highlight that though the risk factor can (advertently or inadvertently) promote hateful content, advertisers can use their economic leverage to require the hosting service to protect against hateful content.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.57 We have clarified that the motivating factors for recorded online hate crime mentioned in the Register of Risks are not exhaustive to reflect the South East Fermanagh Foundation (SEFF) response to the November 2023 Illegal Harms Consultation stating that political stances should be identified as a motivating factor for committing hate offences.

A1.3.58 We have added further evidence demonstrating that the manifestation of online hate can be through several ways (including posts, memes, comments, shared content and a service's 'For You' page).

A1.3.59 We have added prevalence rates of hate-based abuse against the Muslim, Jewish, LGBT+, and disabled, community in the sub-section 'How hate offences manifest online'.

Proposed changes we are not making despite requests to do so

A1.3.60 We have not specifically added politics as a motivating factor for committing a hate crime. However, as noted above, we have addressed the underlying point by clarifying that the motivating factors for hate are not exhaustive.

A1.3.61 We recognise the Institute of Strategic Dialogue's concern around the governance and accountability applied to small high-risk services. We believe the risks posed by these services have been adequately captured in the sub-section 'User base size' in the chapter.

Harassment, stalking, threats and abuse

Summary of stakeholder feedback

A1.3.62 Responses relevant to this chapter of the Register of Risks focused on a wide variety of issues and evidence. In particular, stakeholders provided new evidence linking the relevant offences to a number of existing risk factors, and further insight on how these inter-relate. We received feedback and recommendations regarding areas where further clarity or commentary would help to provide a clearer understanding of how these harms manifest online and the relationships to offline experiences. In particular, we received additional evidence highlighting the distinct risks of harm and impacts for those with protected characteristics. This feedback has been used to update our evidence base linking some risk factors with the relevant offences and provide further clarity around the experience of harm.

A1.3.63 Stakeholders who provided information in their consultation responses relevant to this area included civil society and other independent organisations and individuals (5Rights[99], Centre for Competition Policy[100], Glitch[101], Local Government Association[102], Refuge[103], South East Fermanagh Foundation[104], Name Withheld 3,[105] Dr. Rajan Basra[106], Suzy Lamplugh Trust[107], The Cyber Helpline[108], The Institute for Strategic Dialogue[109]); public bodies (Domestic Abuse Commissioner[110], Scottish Government[111], Victims Commissioner for England and Wales[112]); academics and academic institutions (University College London Gender and Tech Research Group[113]).

A1.3.64 We have decided to make following changes to the relevant Register of Risks chapter.

[99] 5Rights Foundation response to November 2023 Illegal Harms Consultation, p. 8.
[100] Centre for Competition Policy response to November 2023 Illegal Harms Consultation, pp. 1-2.
[101] Glitch response to November 2023 Illegal Harms Consultation, p. 1-3.
[102] Local Government Association response to November 2023 Illegal Harms Consultation, p. 2.
[103] Refuge response to November 2023 Illegal Harms Consultation, p.4.
[104] South East Fermanagh Foundation response to the November 2023 Illegal Harms Consultation, p. 1.
[105] Name Withheld 3 response to November 2023 Illegal Harms Consultation, p. 2.
[106] Dr. Rajan Basra response to the November 2023 Illegal Harms Consultation, p. 1.
[107] Suzy Lamplugh Trust response to the November 2023 Illegal Harms Consultation, pp. 2-4.
[108] The Cyber Helpline response to November 2023 Illegal Harms Consultation, p. 2.
[109] The Institute for Strategic Dialogue response to November 2023 Illegal Harms Consultation, pp. 3-4.
[110] Domestic Abuse Commissioner response to November 2023 Illegal Harms Consultation, pp. 2-5.
[111] Scottish Government response to November 2023 Illegal Harms Consultation, pp. 2-3.
[112] Victims Commissioner for England and Wales response to November 2023 Illegal Harms Consultation, pp. 3-4.
[113] University College London Gender and Tech Research Group response to November 2023 Illegal Harms Consultation, pp. 3-4.

Changes or additions to the evidence base referencing risk factors

A1.3.65 We have added further information and evidence to the risk factor, 'user base demographics', on the specific risks of harm associated with personal characteristics, including gender (women in particular), minority ethnic groups, age and race.

A1.3.66 We have added further information and evidence to the risk factor, 'user base' regarding the experience of women of faith backgrounds.

A1.3.67 We have added further information and evidence to the risk factor, 'user base', regarding the experience of politicians of BAME and marginalised backgrounds.

A1.3.68 We have added the risk factor, 'network recommender systems', with corresponding evidence acknowledging the impact and role these can play in stalking and the proliferation of misogynistic threats/abuse in particular.

Further clarity provided to our conclusions on how evidence is articulated

A1.3.69 We have added further commentary and evidence that provides a more detailed explanation of the link between online activity and how this can facilitate or subsequently manifest in offline risks to victims.

A1.3.70 We have included reference to networked, co-ordinated or large-scale forms of harassment, as distinct from relational harassment, threats and abuse that can be facilitated by the use of online services.

Controlling or coercive behaviour

Summary of stakeholder responses

A1.3.71 Stakeholders provided a variety of new evidence which has been used to update this chapter. In particular, feedback that has allowed us to significantly expand our discussion of how controlling or coercive behaviour ('CCB') manifests online and the link to offline behaviour.

A1.3.72 Stakeholders who provided feedback and sources in their consultation responses relevant to this chapter included civil society and other independent organisations or individuals (Victims' Commissioner for England and Wales[114], Domestic Abuse Commissioner[115], Suzy

[114] Victims' Commissioner for England and Wales response to November 2023 Consultation, p. 5.
[115] Domestic Abuse Commissioner response to November 2023 Consultation, p. 2.

Lamplugh Trust[116], Refuge[117], Glitch[118]); academics and academic institutions (University College London Gender and Technology Research Group[119]). In particular, multiple stakeholders highlighted the importance of linking offline and online behaviour and articulating more clearly important distinctions in experiences and risk in relation to the personal characteristics of victim-survivors.

A1.3.73 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.74 We have added further information on the role that personal characteristics of victims and survivors can play, including disability (User demographics risk factor).

Further clarity provided to our conclusions or how evidence is articulated

A1.3.75 We have expanded the introductory sections of the chapter ('Introduction' and 'How controlling or coercive behaviour manifests online') in regard to how this harm manifests online and the risk of harm experienced by individuals as a result. In particular, the link between offline and online CCB has been discussed in more detail.

A1.3.76 We have added information with regard to wider societal implications of the risk of harm from coercive or controlling behaviour, including the 'chilling' effect on the freedom of speech of women and girls.

Intimate image abuse

Summary of stakeholder responses

A1.3.77 Stakeholders provided new information and evidence that we were able to use to expand on our articulation of how intimate image abuse offences manifest online, including further reference to other forms of intimate image abuse or circumstances in which it has been known to occur. Stakeholders also provided further information on the impact on victims of intimate image abuse. This feedback has enabled us to provide additional detail and expand our evidence base for this chapter.

A1.3.78 Stakeholders who provided information in their consultation responses in relation to this area included civil society and other independent organisations or individuals (Glitch[120],

[116] Suzy Lamplugh Trust response to November 2023 Consultation, p. 2-4.
[117] Refuge response to November 2023 Consultation, pp. 2-3.
[118] Glitch response to November 2023 Consultation, p. 3.
[119] University College London Gender and Technology Research Group response to November 2023 Consultation, p. 3-4.
[120] Glitch response to November 2023 Illegal Harms Consultation, p. 3.

Global Encryption Coalition[121], Refuge[122]); public bodies (Domestic Abuse Commissioner[123], Victims Commissioner[124]); academics and academic institutions (Professor Clare McGlynn of Durham University[125]); and service providers ([✂][126]).

A1.3.79 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.80 We have added further detail and evidence to the risk factor section, 'user base', on the role of intersectionality on risk of being targeted with intimate image abuse and to clarify the way in which both large and small user bases can be relevant to how intimate image abuse offences manifest online.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.81 We have included new references to expand on the issue of intimate image abuse in which synthetic media plays a role, covering the concept of 'deepfakes', generative AI and altered media.

A1.3.82 We have expanded the sub-section 'risk of harm to individuals presented by intimate image abuse' where we explore the impact on victims and survivors and expanded our discussion of the risks of harm to reference wider societal impacts of intimate image abuse.

A1.3.83 We have made clearer reference to the link between intimate image abuse and other harms (e.g. coercive and controlling behaviour).

A1.3.84 We have added further commentary and evidence to the sub-section, 'How intimate image abuse offences manifest online', which demonstrates the prevalence of sextortion.

Extreme pornography

Summary of stakeholder responses

A1.3.85 Stakeholders provided feedback and additional evidence regarding the manifestation of extreme pornography offences and risks of harm that we used to update this chapter.

A1.3.86 Stakeholders who provided feedback and new evidence as part of their consultation responses relevant to this chapter included civil society and other independent

[121] Global Encryption Coalition response to November 2023 Illegal Harms Consultation, pp. 1-2.
[122] Refuge response to November 2023 Consultation, pp. 4.
[123] Domestic Abuse Commissioner response to November 2023 Illegal Harms Consultation, pp. 4-5.
[124] Victims' Commissioner response to November 2023 Illegal Harms Consultation, p. 7.
[125] Prof Clare McGlynn response to November 2023 Illegal Harms Consultation, pp. 3-6.
[126] [✂]

organisations or individuals (CARE[127], Name Withheld 1,[128] SWGfL[129], RSPCA[130], Wildlife and Countryside Link[131], Scottish SPCA[132], Sex Workers Union[133]); academics and academic institutions (Professor Clare McGlynn of Durham University[134]); public bodies (Victims' Commissioner for England and Wales[135]).

A1.3.87 In particular, the responses provided further information on the harm caused by exposure to extreme pornography both to individuals and groups of people, as well as in a broader societal sense. We have been able to significantly expand our discussion on risk and harm.

A1.3.88 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.89 No new risk factors have been added to the chapter since the November 2023 consultation.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.90 We have expanded our discussion on how extreme pornography offences and harm manifests online in response to further evidence and recommendations from stakeholders. In particular, we have expanded our evidence base to consider further how extreme pornography can risk normalising harmful sexual behaviour and violence against women and girls.

A1.3.91 We have included further commentary on the risk of harm from life threatening and high-risk behaviours in response to evidence from stakeholders highlighting the particular risk of harm from content that depicts certain activities (whether real or staged).

[127] CARE response to November 2023 Illegal Harms Consultation, pp. 3-4.
[128] Name Withheld 1 response to November 2023 Illegal Harms Consultation.
[129] SWGfL response to November 2023 Illegal Harms Consultation, p. 3.
[130] RSPCA response to November 2023 Illegal Harms Consultation, p. 1.
[131] Wildlife and Countryside Link response to November 2023 Illegal Harms Consultation, p. 2.
[132] Scottish SPCA response to November 2023 Illegal Harms Consultation, p. 2.
[133] Sex Workers Union response to November 2023 Illegal Harms Consultation, pp. 2-3.
[134] Prof Clare McGlynn response to November 2023 Illegal Harms Consultation, pp. 9-11.
[135] Victims' Commissioner for England and Wales response to November 2023 Illegal Harms Consultation, pp. 4-5.

Sexual exploitation of adults

Summary of stakeholder responses

A1.3.92 Stakeholders provided a range of new sources which have been used to update this chapter.

A1.3.93 Stakeholders who provided feedback and sources in their consultation responses relevant to this chapter included civil society and other independent organisations or individuals (Digital Ventures (Vivastreet)[136], Name Withheld 1[137], Changing Lives[138], Carolina Are[139], Nordic Model Now![140], Global Alliance Against Traffic in Women[141], Refuge)[142]; ([✂][143]).

A1.3.94 Feedback from independent organisations and individuals helped to ensure our description of how these offences manifest online was accurate, and appropriately draws out important nuances between consenting sex workers and those who are coerced or forced into sex work. Particular feedback from law enforcement also focused on ensuring the Register of Risks provides greater clarity on the specific sexual exploitation offences and does not draw overly on the related, but distinct, human trafficking and unlawful immigration offences.

A1.3.95 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.96 We have added further evidence was added in relation to the indicators of potential sexual exploitation online conducted on services which allow 'goods or services to be posted for sale' (Transactions and offers risk factor).

A1.3.97 We have added further evidence for the risk factors, 'user profiles' and 'fake user profiles', to ensure it is focused exclusively on sexual exploitation of adults offences, rather than drawing on evidence related to human trafficking.

A1.3.98 We have included direct reference to Adult Services Websites (ASWs) in response to feedback which demonstrated the critical role these specific sites play in facilitating sexual exploitation offences in the sub-section 'Marketplaces and listings services'.

[136] Digital Ventures (Vivastreet) response to November 2023 Consultation, p. 3.
[137] Name Withheld 1 response to November 2023 Consultation, pp.1-4.
[138] Changing Lives response to November 2023 Consultation, pp. 16-17.
[139] Carolina Are response to November 2023 Consultation, pp. 2-3.
[140] Nordic Model Now! response to November 2023 Consultation, p. 2.
[141] Global Alliance Against Traffic in Women response to November 2023 Consultation, pp. 2-3.
[142] Refuge response to November 2023 Consultation, pp. 2-5.
[143] [✂]

Further clarity provided to our conclusions or how evidence is articulated

A1.3.99 We have updated our discussion of the risks of harm to provide a more nuanced distinction between risks online compared to offline.

A1.3.100 We have sought to provide greater clarity in the sub-section, 'How harms manifest', about the nuances of how these harms are experienced online for different people in different scenarios, and the role of the internet more generally in sex work.

A1.3.101 We have added further analysis on the use of the Sexual Trafficking Indicator Matrix (STIM) in identifying risk factors in the section 'Risk factors: functionalities and recommender systems', particularly on adult services websites.

Human trafficking

Summary of stakeholder responses

A1.3.102 This is a new chapter in the Register of Risks. In our November 2023 consultation it was combined with unlawful immigration. Please see Register of Risks Annex 2 for a discussion of the relevant stakeholder feedback and explanation for this decision.

A1.3.103 Stakeholders provided information and recommendations we used to update the evidence base we had originally used for the combined chapter with unlawful immigration and to draw links between characteristics and the specific human trafficking offences.

A1.3.104 Stakeholders who provided information in their consultation responses relevant to this area included civil society and other independent organisations or individuals (Name Withheld 1[144], Global Alliance Against Traffic in Women[145], Glitch[146], WalkFree[147], Children's Society[148]); public bodies (Scottish Government[149]); academics and academic institutions (Professor Teela Sanders of the University of Leicester[150]); ([✂])[151].

A1.3.105 We have decided to make the following changes to the relevant Register of Risks chapter. These are changes which are in response to, or as a result of, new information provided in response to our November 2023 consultation.

[144] Name Withheld 1 response to November 2023 Illegal Harms Consultation, p. 3.
[145] Global Alliance Against Traffic in Women response to November 2023 Illegal Harms Consultation, pp. 2-3.
[146] Glitch response to November 2023 Illegal Harms Consultation, p. 3.
[147] WalkFree response to November 2023 Illegal Harms Consultation, pp. 2-3.
[148] Children's Society response to May 2024 Protection of Children Consultation, pp. 8-9.
[149] Scottish Government response to November 2023 Illegal Harms Consultation, p. 3.
[150] Prof Teela Sanders response to November 2023 Illegal Harms Consultation, pp. 6-7.
[151] [✂]

Changes or additions to the evidence base referencing risk factors

A1.3.106 Social media services and messaging services (Service type risk factors) have been included, drawing on existing and new evidence provided by stakeholders.

A1.3.107 We have added further evidence to the risk factor, 'marketplace and listings services', strengthening our existing analysis.

A1.3.108 We have added to our commentary and conclusions regarding the role of the risk factor, 'ephemeral messaging', in cases of exploitation.

A1.3.109 We added 'posting content (images, videos, hashtags, emojis)' and 'livestreaming' (user communications risk factors) in response to evidence provided by stakeholders that demonstrated the role these features have played in human trafficking cases.

A1.3.110 We added 'posting or sending location information' (user communications risk factors) as a risk factor in response to new information received from stakeholders.

A1.3.111 'User groups and connections' (user networking risk factor) has also been included as a result of more comprehensive evidence provided by stakeholders relating this to human trafficking offences.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.112 We have updated the 'how human trafficking offences manifest online' sub-section of the chapter to provide greater clarity with respect to the specific relevant offences and how human trafficking manifests online. In particular, we have provided further detail on the different forms of human trafficking offences (namely, sexual exploitation and abuse, criminal exploitation including county lines, forced labour and exploitation, and trafficking offences and migrant victims).

A1.3.113 As part of splitting out the chapter from unlawful immigration, we have removed sources that were related only to unlawful immigration and not human trafficking.

Unlawful immigration

Summary of stakeholder responses

A1.3.114 This is a new chapter in the Register of Risks. In our November 2023 consultation, it was combined with human trafficking. Please see Register of Risks Annex 2 for a discussion of the relevant stakeholder feedback and explanation for this decision.

A1.3.115 A number of stakeholders provided information and recommendations we used to provide greater clarity on how this harm manifests, particularly as a separate kind of illegal harm to human trafficking.

A1.3.116 Stakeholders who provided information in their consultation responses relevant to this area included civil society and other independent organisations or individuals (Global Alliance Against Traffic in Women[152], Glitch[153], Name Withheld 1[154]); public bodies (Scottish Government[155]); ([✂][156]).

A1.3.117 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.118 We have added further evidence to the risk factor, 'user base size', with evidence from stakeholders demonstrating how unlawful immigration content spreads widely on services with large userbases.

A1.3.119 We have expanded our discussion of the risk factor, 'user base demographics', highlighting the importance of other personal characteristics such as socio-economic or mental health factors.

A1.3.120 We have added further evidence to our risk factors, 'user profiles, fake user profiles, and anonymous user profiles', to show how perpetrators can abuse features from a profiles or an account to advertise unlawful immigration services.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.121 We have updated the early sections of the chapter to provide greater clarity in regard to the specific relevant offences and how unlawful immigration manifests online, and the risk to UK citizens.

A1.3.122 As part of splitting out the chapter from human trafficking, we have removed sources that were related only to human trafficking and not unlawful immigration.

Fraud and financial services offences

Summary of stakeholder responses

A1.3.123 Stakeholders provided a range of new sources which have been used to update this chapter. In particular, stakeholders shared new evidence which has allowed us to

[152] Global Alliance Against Traffic in Women response to November 2023 Illegal Harms Consultation, pp. 2-3.
[153] Glitch response to November 2023 Illegal Harms Consultation, p. 3.
[154] Name Withheld 1 response to November 2023 Illegal Harms Consultation, p. 3.
[155] Scottish Government response to November 2023 Illegal Harms Consultation, p. 3.
[156] [✂]

strengthen the evidence base linking various risk factors with fraud, as well as providing further evidence in relation to the different victims of online fraud.

A1.3.124 Stakeholders who provided feedback and new evidence as part of their consultation responses relevant to this chapter included civil society and other independent organisations or individuals (Alliance to Counter Crime Online[157], Global Encryption Coalition[158], Innovate Finance[159], Integrity Institute[160], JobsAware[161], Lloyds Banking Group[162], Which?[163], [✂][164], Revolut[165], StopScams[166], TSB Bank[167], UK Finance[168], Cifas[169]); public bodies ([✂])[170], National Trading Standards eCrime Team[171], Victims Commissioner for England and Wales[172]); ([✂][173]); service providers (Google[174], Mid Size Platform Group[175]).

A1.3.125 We have decided to make the following changes to relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.126 We have updated statistics in relation to the volume of fraud, and cyber-enabled fraud, in the UK within the introduction to this chapter.

A1.3.127 We have included more detail on the role that different personal characteristics can play in increasing the likelihood that any one person is a victim of different types of fraud (User base demographics).

A1.3.128 We have included further detail and additional evidence to the risk factors, 'user profiles' and 'fake user profiles', to show how they are utilised in the facilitation of fraud online.

A1.3.129 We have added further evidence to the risk factor, user groups, drawing on evidence that highlighted the specific role of online 'influencers'.

---

[157] Alliance to Counter Crime Online response to November 2023 Illegal Harms Consultation, p. 3.
[158] Global Encryption Coalition response to November 2023 Illegal Harms Consultation, pp. 1-2.
[159] Innovate Finance response to November 2023 Illegal Harms Consultation, pp. 3-4.
[160] Integrity Institute response to November 2023 Illegal Harms Consultation, p. 1.
[161] JobsAware response to November 2023 Illegal Harms Consultation, pp. 2-3.
[162] Lloyds Banking Group response to November 2023 Illegal Harms Consultation, pp. 3-4.
[163] Which? response to November 2023 Illegal Harms Consultation, pp. 3-4.
[164] [✂]
[165] Revolut response to November 2023 Illegal Harms Consultation, pp. 2-6.
[166] StopScams response to November 2023 Illegal Harms Consultation, pp. 3-4.
[167] TSB Bank response to November 2023 Illegal Harms Consultation, p. 2.
[168] UK Finance response to November 2023 Illegal Harms Consultation, pp. 1-17
[169] Cifas response to November 2023 Illegal Harms Consultation, p. 2.
[170] [✂]
[171] National Trading Standards eCrime Team response to November 2023 Illegal Harms Consultation, p. 1.
[172] Victims Commissioner for England and Wales response to November 2023 Illegal Harms Consultation, p. 8.
[173] [✂]
[174] Google response to November 2023 Illegal Harms Consultation, pp. 9-10.
[175] Mid Size Platform Group response to November 2023 Illegal Harms Consultation, p. 2.

A1.3.130 We have added the risk factor, 'hyperlinking' (user communications risk factor) for consideration, in response to evidence demonstrating that they play an important role in certain types of fraud.

A1.3.131 We have added additional evidence to the risk factor, 'posting goods or services for sale', (Transactions and offers risk factor) due to how important a functionality this is for committing purchase scams, drawing on additional evidence shared by stakeholders.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.132 We have provided more examples and descriptions of the different types of fraud that occur online, and provided additional detail about the link between fraud and proceeds of crime offences.

A1.3.133 We have provided further clarity regarding the role of revenue models and the risks of fraud (paragraph 11.97), in particular the role of transaction fees and 'pay to promote' functionality that can provide opportunities for fraudsters to promote fraudulent content.

Proceeds of crime

Summary of stakeholder responses

A1.3.134 Stakeholders provided new information and evidence we have used to add further detail to our description of how 'proceeds of crime offences manifest online'.

A1.3.135 Stakeholders who provided feedback and new evidence as part of their consultation responses relevant to this chapter included civil society and other independent organisations or individuals (The Children's Society[176], TSB Bank[177], UK Finance[178], Cifas[179]).

A1.3.136 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.137 We have included the risk factor, 'gaming services', in response to new evidence provided at consultation that pointed to multiple reports of these services being used by perpetrators to identify and initiate contact with potential money mules, including children.

[176] The Children's Society response to May 2024 Protection of Children Consultation, p. 11.
[177] TSB Bank response to November 2023 Consultation, p. 2.
[178] UK Finance response to November 2023 Consultation, pp. 1-17.
[179] Cifas response to November 2023 Consultation, p. 2.

A1.3.138 We have added the risk factor, 'user groups', as a result of evidence that these have been used to facilitate proceeds of crime offences, particularly through the sharing of information or guidance on how to commit certain offences.

A1.3.139 We have updated statistics on the instances of money muling.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.140 We have added additional detail on the link between proceeds of crime offences and other related criminal activity (such as drug trafficking, human trafficking and cyber-crimes). We have also added additional detail highlighting the specifics risks to children who are groomed into illegal activity such as money laundering.

Drugs and psychoactive substances

Summary of stakeholder responses

A1.3.141 Responses relevant to this chapter of the Register of Risks focused predominantly on the provision of new evidence that supported existing conclusions, as well as providing additional examples of how drugs offences manifest online, and more specificity about the kinds of illegal substances that are being bought and sold online in the UK. These responses enabled us to update our evidence base linking some risk factors with drugs and psychoactive substances offences.

A1.3.142 Stakeholders who provided new information in this area included academics (Ashly Fuller of University College London)[180] and law enforcement (Association of Police and Crime Commissioners)[181].

A1.3.143 We have decided to make following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.144 We have added the risk factor, 'Services where users can post or send location information', with corresponding evidence that shows its link to the sale of illicit drugs.

A1.3.145 We have added the risk factor, 'Dating services', in response to evidence highlighting that dating and hookup sites have been used extensively for the sale of substances, particularly those which are catered to narrow audiences.

A1.3.146 We have added the risk factor, 'Discussion forum or chat room service', in response to evidence provided in consultation response that highlighted the use of these services in

[180] Ashly Fuller response to November 2023 Illegal Harms Consultation, p. 9.
[181] Association of Police and Crime Commissioners response to November 2023 Illegal Harms Consultation, p. 3.

the sale of illicit drugs online, as well as new evidence found exploring this issue in other European countries.

A1.3.147 We have added further evidence from law enforcement stakeholders, and from more recently published evidence, supporting conclusions about a number of risk factors, particularly in relation to messaging services and ephemeral messaging.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.148 We have updated our commentary on the particular risk posed by the increased use, and access to online, of new opioids (including nitazenes).

A1.3.149 We have included further information on the relationship between online drugs sales and other forms of criminal activity, including crimes with an online component (such as the non-consensual recording and sharing online of intimate images).

Proposed changes we are not making despite requests from stakeholders to do so

A1.3.150 We do not agree that marketplaces should be excluded from the risk factors identified in this chapter given the evidence which points to their use in the sale of, or offers to sell, drugs and psychoactive substances.[182] However, we recognise that there are a variety of types of marketplace and listing services and individual services will need to consider their own circumstances when assessing and mitigating risks.

Firearms, knives and other weapons

Summary of stakeholder responses

A1.3.151 Responses relevant to this chapter of the Register of Risks focused predominantly on the provision of new evidence that supported existing conclusions, as well as providing additional evidence. Stakeholders who provided new information in this area included civil society organisations (5Rights)[183]; [✂[184]].

A1.3.152 We have decided to make the following changes to the Register of Risks chapter 'Firearms, knives and other weapons offences' for publication.

[182] A small number of services raised this concern, citing their existing governance processes and mitigations, as well as the fact their service is focused on selling one particular type of product or service that would reasonably be considered to reduce the risk of their particular service being misused for the sale of illegal items.
[183] 5Rights response to the November 2023 Illegal Harms Consultation, p. 4.
[184] [✂]

Changes or additions to the evidence base referencing risk factors

A1.3.153 We have included further evidence to the risk factor, 'online marketplaces', for the sale of / offers to sell illegal weapons.

A1.3.154 We have included further evidence to the risk factor, 'Services where users can post or send content anonymously, including without an account', supporting conclusions about the role of anonymity online in the sale of / offer for sale of illegal weapons.

A1.3.155 We have included the risk factor, 'Social media', to evidence highlighting the role these services can play in facilitating the sale of illegal weapons, alongside the wider issue of the visibility and glamourisation of weapons and knife crime in particular.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.156 Clarifying the specific types of weapons identified as being accessible through e-commerce sites in cited evidence.

A1.3.157 Clarifying that the minimum cost of a potentially illegal firearm online we had quoted from research exploring whether illegal weapons could be purchased via online marketplaces was likely to be referring to a component of a firearm, rather than a fully functioning firearm (paragraph 14.33).

A1.3.158 We have moved references of 3D printed firearms to the Terrorism chapter. This is because the provision of blueprints or instructions which allow another person to 3D print ('make') a firearm would be considered illegal content in relation to offences relating to the provision of instructions or training in the making or use of firearms. Refer to the Terrorism chapter of the ICJG for further details.

Encouraging or assisting suicide (or attempted suicide)

Summary of stakeholder responses

A1.3.159 In our draft Register of Risks published in our November 2023 Consultation this chapter also included the non-priority offence of encouraging or assisting serious self-harm as at the time it was not clear whether this would be a priority or non-priority offence in the Act. Please see Register of Risks Annex 2 for an explanation for this decision. Stakeholder feedback was provided in relation to encouraging and assisting both suicide and self-harm and was used to update all relevant chapters of the Register of Risks.

A1.3.160 Stakeholders who provided feedback and sources in their consultation responses relevant to this area included civil society and other independent organisations or individuals

(CarefulAI[185], Samaritans[186], NSPCC[187], Zevo Health[188], Canadian Centre for Child Protection[189], 5Rights Foundation[190], Mental Health Foundation[191], Glitch[192], Molly Rose Foundation[193], TalkLife,[194], OSAN[195]); public bodies (Scottish Government[196]); service providers ([✂][197]).

A1.3.161 Stakeholders predominantly provided additional information and evidence related to the impacts on users who encounter this kind of content online and wider trends within society which provide useful context for these experiences, for example, issues such as the 'contagion effect'. Stakeholders also provided additional information on the role of personal characteristics (such as age and gender) in increasing risks of both encountering and being impacted by this kind of content.

A1.3.162 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.163 We have included further evidence on the role of services with small user bases ('User base size' risk factor).

A1.3.164 Using new evidence and recommendations from stakeholder responses, we have explored in greater detail the role of personal characteristics in encountering potentially illegal content and the risk of harm that results, including demographic characteristics (such as age and sexual orientation), as well as personal circumstances including people's mental health ('User demographics' risk factor).

A1.3.165 We have included further evidence provided by stakeholders in relation to users being able to save content that has been posted on U2U services (the 'User communication' risk factor 'Posting content (text, images, videos)'.

A1.3.166 The evidence base around the role of 'content recommender systems' (Recommender systems risk factor) has been expanded, alongside further discussion of U2U business models in which these and other functionalities play an important role.

---

[185] CarefulAI response to November 2023 Consultation, p. 2.
[186] Samaritans response to November 2023 Consultation, p. 2.
[187] NSPCC response to November 2023 Consultation, pp. 4-5.
[188] Zevo Health response to November 2023 Consultation, p. 2.
[189] Canadian Centre for Child Protection response to November 2023 Consultation, pp. 1-3.
[190] 5Rights Foundation response to November 2023 Consultation, pp. 4-6.
[191] Mental Health Foundation response to November 2023 Consultation p. 1.
[192] Glitch response to November 2023 Consultation, p. 2.
[193] Molly Rose Foundation response to November 2023 Consultation, pp. 8-39.
[194] TalkLife response to November 2023 Consultation, pp. 1-3.
[195] OSAN response to November 2023 Consultation, p. 37.
[196] Scottish Government response to November 2023 Consultation, p. 3.
[197] [✂]

A1.3.167 We have added evidence the 'business models and commercial profile' section of the chapter, drawing on evidence provided by stakeholders that highlights, in particular, the link between recommender systems and business models. Further information on this broader concept has also been included in the Introduction to the Register of Risks.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.168 We have updated the sub-section 'how encouraging or assisting suicide manifests online' in regard to the risk of harm experienced by individuals as a result of exposure to potentially illegal content, in particular including further commentary regarding the normalisation of certain behaviours in response to exposure.

A1.3.169 As part of splitting out the chapter from self-harm, we have removed sources from this chapter that exclusively focused on content related to self-harm. This evidence can be found in the new chapter 'Encouraging or assisting serious self-harm'.

Foreign interference

Summary of stakeholder responses

A1.3.170 Stakeholders provided input and a range of new evidence which have been used to update this chapter. In their responses, stakeholders provided further commentary and additional sources of information regarding the wider context and broader offline forms of conduct that are linked to the offence online, as well as highlighted the risks of generative AI. This chapter has been expanded significantly since the draft published in the November 2023 Illegal Harms consultation due, in large part, to the growing evidence base in this area and developments in generative AI technology.

A1.3.171 Stakeholders who provided information in their consultation responses relevant to this area included civil society and other independent organisations or individuals (Glitch[198], Logically[199], Global Disinformation Index[200]); academics and academic institutions (Swansea University[201]).

A1.3.172 We have decided to make following changes to the relevant Register of Risks chapter.

[198] Glitch response to November 2023 Illegal Harms Consultation, p. 5.
[199] Logically response to November 2023 Illegal Harms Consultation.
[200] Global Disinformation Index response to November 2023 Illegal Harms Consultation, p. 3.
[201] Swansea University response to November 2023 Illegal Harms Consultation, pp. 2-3.

Changes or additions to the evidence base referencing risk factors

A1.3.173 We have added further detail and evidence to the risk factor sub-section, 'user base'. This new evidence provides additional detail on the experiences and risk of exposure to, and potential for harm to occur as a result of, foreign interference for women, minority ethnic groups, certain sexualities and religion, as well as experiences of individuals with multiple characteristics.

A1.3.174 We have added further evidence to the risk factor, 'user base demographics', regarding individuals with intersectional identities being at an increased risk of foreign interference.

A1.3.175 We have added further evidence to the risk factor, 'fake user profiles', which cites real world cases where fake user profiles have been used by one nation to undermine another.

A1.3.176 We have added further evidence to the risk factor, 'posting content', to highlight cases where generative AI content has been used in foreign interference operations.

A1.3.177 We have added further evidence to the risk factor, 'hyperlinking', regarding a case where hyperlinks were used to influence elections in the U.S.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.178 We have updated our commentary and added further evidence to the section 'how foreign interference offence can manifest online', including a high-level overview of the role of Generative AI in facilitating foreign interference.

A1.3.179 We have updated our commentary and added further evidence to the sub-section 'risk of harm to individuals presented by foreign interference offences', in relation to societal harms, threats to public health, threats to information systems and democratic processes, threats to women and people of colour's participation in public life.

Animal cruelty

Summary of stakeholder responses

A1.3.180 Stakeholders provided a range of additional sources and recommendations in relation to the Animal Cruelty offence that we have been able to use to make updates to the Register of Risks. In particular, stakeholders provided further examples of how this offence could manifest online.

A1.3.181 Stakeholders who provided feedback and new evidence as part of their consultation responses relevant to this chapter included civil society and other independent

organisations or individuals ([✂][202]; Name Withheld 1 (August 2024)][203]; Battersea Dogs and Cats Home[204]; Blue Cross[205]; Born Free Foundation[206]; Cats Protection[207]; Dogs Trust[208]; Humane Society[209]; International Cat Care[210]; OSAN[211]; RSPCA[212]; Scottish SPCA[213]; Social Media Animal Cruelty Coalition[214]; Wildlife and Countryside Link[215]; Southwest Grid for Learning[216]) and service providers (Google[217], Mid Size Platform Group[218]).

A1.3.182 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.183 We have added examples of online content encouraging mutilations (for example, ear cropping) and of hunting to the sub-sections 'Social media and video-sharing services' and 'Messaging services, discussion forums and chat room services'.

A1.3.184 We have added a sentence to paragraph 17.33 to acknowledge more clearly that users could provide hyperlinks to direct messaging services to encourage animal cruelty or conspire with others.

A1.3.185 We have included a sentence about livestreaming potentially being used within animal torture networks, and a media source evidencing this in a footnote (paragraph 17.56).

Further clarity provided to our conclusions or how evidence is articulated

A1.3.186 We have made slight changes throughout the chapter to provide more clarity to what constitutes the relevant offence, especially as regards the 'commissioning of offences'.

A1.3.187 We have added paragraph 17.18 to explain that the Register of Risks may consider a broader evidence base than would necessarily meet the threshold for the priority offence. This is because we acknowledge that some pre-recorded content could help to normalise

---

[202] [✂]
[203] Name Withheld 1 response to August 2024 Consultation on Torture and Animal Cruelty.
[204] Battersea Dogs and Cats Home response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-9.
[205] Blue Cross response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-5.
[206] Born Free Foundation response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-3.
[207] Cats Protection response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-7.
[208] Dogs Trust response to August 2024 Consultation on Torture and Animal Cruelty, p. 4.
[209] Humane Society response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-3.
[210] International Cat Care response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-5.
[211] OSAN response to August 2024 Consultation on Torture and Animal Cruelty, p. 2.
[212] RSPCA response to August 2024 Consultation on Torture and Animal Cruelty, pp. 1-4.
[213] Scottish SPCA response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-5.
[214] Social Media Animal Cruelty Coalition response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-3.
[215] Wildlife and Countryside Link response to August 2024 Consultation on Torture and Animal Cruelty, p. 2.
[216] Southwest Grid for Learning response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-4.
[217] Google response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-3.
[218] Mid Size Platform Group response to August 2024 Consultation on Torture and Animal Cruelty. p. 3.

certain behaviours, which in turn could lead to acts which *do* constitute the priority offence.

A1.3.188 We have added some clarity relating to 'fake rescue' content:

- We altered the definition of this content so it relates to 'harmful situations', of which predator/prey scenarios are just one example rather than the whole genre;
- We added an additional footnote under the 'user connections' risk factor, drawing a link between building large user followings and a potential sense of legitimacy. This is mentioned in the Fraud and financial services offences chapter of the Register of Risks.

A1.3.189 We re-framed 'direct messaging and encrypted messaging' (user communication risk factor), noting that these functionalities could *contribute* to the risk of 'messaging services' (service type risk factor) being used for the offence. We also moved evidence of animal torture networks using polling functions from 'direct messaging' to 'user groups and group messaging' (user networking risk factor) where we believe it is more relevant.

A1.3.190 We have included some additional clarity to how bad actors could take advantage of content recommender systems to commit the animal cruelty offence (paragraph 17.71)

Proposed changes we are not making despite requests from stakeholders to do so

A1.3.191 We have not added 'gaming' and 'moderation processes' as risk factors in the Register of Risks. We reviewed the evidence submitted by stakeholders about the link between video games and attitudes towards animals, but do not feel it is a relevant risk factor for the ability to use online services to encourage, assist or conspire to commit animal cruelty. Similarly, while we recognise that bad actors posting illegal content may seek to evade moderation systems, we do not feel there are strong, evidenced links between this particular risk factor and the ability to commit the animal cruelty priority offence.

A1.3.192 Some stakeholders called for certain risk factors that are already in the Register of Risks to be made 'key' for animal cruelty, such as 'anonymous posting' (anonymity is covered in the Risk factors: functionalities and recommender systems section). We consider risk factors to be key where they are strongly linked to the illegal harm, based on the evidence we have assessed - see our Risk Profiles for more information.

A1.3.193 We have also not included 'hyperlinking to bestiality content hosted on user-to-user pornography services or on other types of services' as a risk factor. Bestiality falls under the extreme pornography offence, for which user-to-user pornography services are a key service-type risk factor.

A1.3.194 We have not expanded our existing commentary and evidence base on the links between animal abuse and future offending or other forms of abuse, such as child abuse and

domestic abuse. There is, to date, very little research studying the impact of viewing of animal cruelty content *online*.

A1.3.195 We have not included advertisements for free dogs as evidence for the use of marketplace or listing services being a risk factor for the offence. Even if some people intending on committing acts of animal cruelty may search for free animals online, posting an advertisement is not illegal, unless it could be proven that the poster was intending to facilitate animal cruelty. In addition, these advertisements do not cause demonstrable harm to UK users.

A1.3.196 We did not think it relevant to include what one stakeholder called 'casual cruelty', such as social media trends that inadvertently cause distress to animals or advertisements for inappropriate hutches. This would not constitute the priority offence, and we know of no evidence that bad actors share this type of content with the intent to normalise certain behaviours or encourage extreme acts of cruelty.

Epilepsy trolling

Summary of stakeholder responses

A1.3.197 Stakeholders who provided feedback as part of their consultation responses relevant to epilepsy trolling included civil society and other independent organisations or individuals (5Rights Foundation[219], SafeCast (3)[220]); public bodies (Scottish Government[221]); and service providers (LinkedIn[222]).

A1.3.198 Where epilepsy trolling was mentioned, responses were focused on providing feedback on the risk assessment process, the Illegal Content Judgements Guidance (ICJG) and the suitability of Codes measures. In the absence of additional evidence or recommendations for how the draft Register of Risks could be amended within Consultation responses, and any further additional published evidence identified, we have not made any changes in the Register of Risks with regards to Epilepsy Trolling.

Cyberflashing

Summary of stakeholder responses

A1.3.199 Stakeholders provided information and recommendations which we have used to expand on our discussion of how the cyberflashing offence manifests online, including more

[219] 5Rights Foundation response to November 2023 Illegal Harms Consultation, pp. 4-7.
[220] SafeCast (3) response to November 2023 Illegal Harms Consultation, pp. 6-7.
[221] Scottish Government response to November 2023 Illegal Harms Consultation, p. 13.
[222] LinkedIn response to November 2023 Illegal Harms Consultation, p. 20.

nuanced discussion of the mindset of those sending intimate images and the impacts this behaviour can have on victims.

A1.3.200 Stakeholders who provided information in their consultation responses relevant to this area included civil society and other independent organisations and individuals (Glitch[223], Refuge[224]); public bodies (Domestic Abuse Commissioner[225], Victims Commissioner[226]); academics and academic institutions (Professor Clare McGlynn of Durham University[227]).

A1.3.201 We have decided to make the following changes to the relevant Register of Risks chapter.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.202 We have expanded our discussion in regard to user base demographics and the relative risk of harm for people with different characteristics.

Encouraging or assisting serious self-harm

Summary of stakeholder responses

A1.3.203 In our draft Register of Risks published in our November 2023 Consultation this harm was grouped together with the priority offence of 'encouraging or assisting suicide' as, at the time, it was not clear that 'encouraging or assisting serious self-harm' would be a non-priority offence in the Act. Please see Register of Risks Annex 2 for an explanation for the decision to separate these harms. Stakeholder feedback was provided in relation to encouraging and assisting both suicide and self-harm and was used to update all relevant chapters of the Register of Risks.

A1.3.204 Stakeholders provided valuable input and a range of new sources which have been used to update this chapter. Stakeholder feedback was provided in relation to encouraging and assisting both suicide and self-harm and was used to update all relevant chapters of the Register of Risks.

A1.3.205 Stakeholders who provided feedback and new evidence as part of their consultation responses relevant to this chapter included civil society and other independent organisations or individuals (5Rights Foundation[228], Canadian Centre for Child Protection

[223] Glitch response to November 2023 Illegal Harms Consultation, p. 3.
[224] Refuge response to November 2023 Consultation, p. 5.
[225] Domestic Abuse Commissioner response to November 2023 Illegal Harms Consultation, pp. 2-5.
[226] Victims Commissioner response to November 2023 Illegal Harms Consultation, p. 8.
[227] Prof Clare McGlynn response to November 2023 Illegal Harms Consultation, pp. 6-8.
[228] 5Rights Foundation response to November 2023 Illegal Harms Consultation, pp. 4-6.

(C3P)[229], CarefulAI[230], Glitch[231], Mental Health Foundation[232], Molly Rose Foundation,[233] NSPCC[234], Samaritans,[235] Online Safety Act Network[236], TalkLife[237], Zevo Health[238]); public bodies (Scottish Government[239]); service providers ([✂[240]]).

A1.3.206 The following changes of each type have been made to the Register of Risks chapter 'Encouraging or assisting serious self-harm' for publication.

Changes or additions to the evidence base referencing risk factors

A1.3.207 We have added further evidence provided by stakeholders to the risk factor, 'user base demographics'. In particular, highlighting the relatively greater risk of encountering this content among younger users, female users, and those with neurological and psychological conditions.

A1.3.208 We have added further evidence provided by stakeholders to the risk factor, 'commenting on content'.

A1.3.209 We have added further evidence from stakeholders and updated our commentary to the risk factor, 'Recommender systems', to include additional reference to the prompts for recommended content that users can encounter.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.210 We have provided additional commentary and evidence to the sub-section 'How encouraging or assisting serious self-harm manifests online', including additional evidence pointing to the risks of normalisation of self-injurious behaviour posed by exposure to this content.

---

[229] Canadian Centre for Child Protection (C3P) response to November 2023 Illegal Harms Consultation, pp. 33.
[230] CarefulAI response to November 2023 Illegal Harms Consultation, p. 1.
[231] Glitch response to November 2023 Illegal Harms Consultation, p. 2.
[232] Mental Health Foundation response to November 2023 Illegal Harms Consultation, p. 1.
[233] Molly Rose Foundation response to November 2023 Illegal Harms Consultation, pp. 8, 12-27, 38-39.
[234] NSPCC response to November 2023 Illegal Harms Consultation, pp. 4-5.
[235] Samaritans response to November 2023 Illegal Harms Consultation, p. 2.
[236] Online Safety Act Network response to November 2023 Illegal Harms Consultation, pp. 15, 46.
[237] TalkLife response to November 2023 Illegal Harms Consultation, pp. 1-3.
[238] Zevo Health response to November 2023 Illegal Harms Consultation, p. 2.
[239] Scottish Government response to November 2023 Illegal Harms Consultation, p. 3.
[240] [✂]

False communications

Summary of stakeholder feedback

A1.3.211 Stakeholders provided additional evidence regarding false communication online, in particular the additional risks posed by AI-generated 'deepfake' content.

A1.3.212 Stakeholders who provided feedback and new evidence as part of their consultation responses relevant to this chapter included civil society and other independent organisations or individuals (Glitch[241], Logically,[242] Global Disinformation Index[243]); academics and academic institutions (Swansea University[244]).

A1.3.213 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.214 We have added further evidence to the risk factor, 'posting content', due to a series of new high-profile examples and recent evidence provided by stakeholders and published since the November 2023 Consultation.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.215 We have expanded on our commentary regarding the role of generative AI in enhancing the risks posed by false communications.

Proposed changes we are not making despite requests from stakeholders to do so

A1.3.216 We received a response stating that there should be an exploration of intersectional harms in this area. However, there is limited evidence on the role of intersectionality in propagating the offence. Therefore, we have decided to not update our chapter with regards to this area.

[241] Glitch response to November 2023 Illegal Harms Consultation, p. 2.
[242] Logically response to November 2023 Illegal Harms Consultation. pp. 4-9.
[243] Global Disinformation Index response to November 2023 Illegal Harms Consultation, pp. 1-2.
[244] Swansea University response to November 2023 Illegal Harms Consultation, pp. 1-4.

Obscene content showing torture of humans and animals (the s.127(1) offence)

Summary of stakeholder feedback

A1.3.217 Stakeholders who provided feedback relevant to this chapter as part of their consultation responses included civil society and other independent organisations or individuals (Name Withheld 1 (August 2024)[245]; Battersea Dogs and Cats Home[246]; Blue Cross[247]; Born Free Foundation[248]; Cats Protection[249]; Dogs Trust[250]; [✂[251]]; Humane Society[252]; International Cat Care[253]; OSAN[254]; RSPCA[255]; Scottish SPCA[256]; Social Media Animal Cruelty Coalition[257]; Wildlife and Countryside Link[258]; Southwest Grid for Learning[259]) and service providers (Google[260], Mid Size platform Group[261])

A1.3.218 We have decided to make the following changes to the relevant Register of Risks chapter.

Further clarity provided to our conclusions or how evidence is articulated

A1.3.219 We have rewritten the commentary on the evidence of gender as a risk factor for user harm from viewing explicit sexual or violent content, for the purpose of clarity.

A1.3.220 We have added a cross-reference (paragraph 22.43) to the chapter on the animal cruelty offence in relation to hyperlinking to messaging and file-sharing services as a risk factor.

[245] Name Withheld 1 response to August 2024 Consultation on Torture and Animal Cruelty, pp. 1-3.
[246] Battersea Dogs and Cats Home Home response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-9.
[247] Blue Cross response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-5.
[248] Born Free Foundation response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-3.
[249] Cats Protection response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-7.
[250] Dogs Trust response to August 2024 Consultation on Torture and Animal Cruelty, p. 4.
[251] [✂]
[252] Humane Society response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-3.
[253] International Cat Care response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-5.
[254] OSAN response to August 2024 Consultation on Torture and Animal Cruelty, p. 2.
[255] RSPCA response to August 2024 Consultation on Torture and Animal Cruelty, pp. 1-4.
[256] Scottish SPCA response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-5.
[257] Social Media Animal Cruelty Coalition response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-3.
[258] Wildlife and Countryside Link response to August 2024 Consultation on Torture and Animal Cruelty, p. 2.
[259] Southwest Grid for Learning response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-4.
[260] Google response to August 2024 Consultation on Torture and Animal Cruelty, pp. 2-3.
[261] Mid Size Platform Group response to August 2024 Consultation on Torture and Animal Cruelty. p. 3.

Search services

Summary of stakeholder responses

A1.3.221 Stakeholders provided a range of new sources of evidence and recommendations which have been used to update this chapter.

A1.3.222 Stakeholders who provided feedback and new evidence as part of their consultation responses relevant to this chapter included civil society and other independent organisations or individuals (Refuge)[262]; service providers (Google[263], Mid Size Platform Group,[264] Skyscanner[265]).

A1.3.223 We have decided to make the following changes to the relevant Register of Risks chapter.

Changes or additions to the evidence base referencing risk factors

A1.3.224 We have added further evidence in relation to how the risk factor 'general search services', enables users to access various types of illegal content (including CSAM and non-consensual intimate images) (Service types risk factor).

A1.3.225 We have updated our commentary of how the risk factor 'user base size' can play a role in risks of harm, in response to questions about how both large and small user bases can both increase risk (User base risk factor).

A1.3.226 We have added further evidence to the sub-section 'Search query inputs' by adding evidence that those who actively seek out illegal content may be more vulnerable to experience or cause harm as a result.

A1.3.227 We have added further evidence to the risk factor, 'image search', in relation to how reverse image search in particular can be used to carry out offences such as harassment, coercive controlling behaviour and intimate image abuse (Search query inputs risk factor).

Further clarity provided to our conclusions or how evidence is articulated

A1.3.228 We have added further commentary to the sub-section 'how harm manifests on search services', in particular focusing on how harm occurs as result of encountering illegal content via search services and how other chapters of the Register of Risks on U2U services can provide useful information.

[262] Refuge response to November 2023 Illegal Harms Consultation, p. 5.
[263] Google response to November 2023 Illegal Harms Consultation, pp. 5-9.
[264] Mid Size Platform Group response to November 2023 Illegal Harms Consultation, p. 3.
[265] Skyscanner response to November 2023 Illegal Harms Consultation, p. 3.

A1.4 Risk Profiles

Stakeholder responses by theme

Support for our approach

A1.4.1 Betting and Gaming Council, Nexus, Oxford Disinformation and Extremism Lab, and Segregated Payments Ltd agreed with and supported our proposals.[266]

A1.4.2 Logically believed that the Risk Profiles were sufficiently clear on how the foreign interference illegal harm can manifest and how the risks of the associated offences being committed or facilitated on services should be assessed.[267] SPRITE+ also told us that they found the Risk Profiles clear, accessible and proportionate.[268]

A1.4.3 Vinted stated that our four-step risk assessment process and Risk Profiles appear be in line with their current internal risk management approach.[269]

A1.4.4 Mid Size Platform Group were supportive of our decision not to include generative AI in the first iteration of the Risk Profiles, given the lack of available evidence currently.[270]

A1.4.5 Microsoft welcomed our efforts to keep both the Illegal Harms and Children's Risk Profiles as consistent as possible.[271]

A1.4.6 UKIE considered the step-by-step process outlined for consulting the Children's U2U Risk Profile was clear and methodical. This included the questions we ask so that service providers can identify specific risk factors and the inclusion of a glossary to help them interpret these risk factors. This, it said, ensures that all service providers, regardless of their size or resources, can effectively use the Risk Profiles.[272]

Feedback on the format of the Risk Profiles

A1.4.7 Protection Group International stated that the Risk Profiles needed to be presented in a clearer way that enabled easier identification of all relevant offences.[273] Furthermore, an individual did not agree that the Risk Profiles were clear and suggested that additional

[266] Betting and Gaming Council response to November 2023 Illegal Harms Consultation, p.4; Nexus response to November 2023 Illegal Harms Consultation, pp.5-6; Oxford Disinformation and Extremism Lab response to November 2023 Illegal Harms Consultation, pp.3-4; Segregated Payments Ltd response to November 2023 Illegal Harms Consultation, p.4.
[267] Logically response to November 2023 Illegal Harms Consultation, p.12.
[268] SPRITE+ response to November 2023 Illegal Harms Consultation, p.7.
[269] Vinted response to November 2023 Illegal Harms Consultation, p.4.
[270] Mid Size Platform Group response to November 2023 Illegal Harms Consultation, p.4.
[271] Microsoft response to May 2024 Consultation on Protecting Children from Harms Online, p.6.
[272] UK Interactive Entertainment Association response to May 2024 Consultation on Protecting Children from Harms Online, p.21-25. Our approach to the Illegal Harms Risk Profiles is the same.
[273] Protection Group International to November 2023 Illegal Harms Consultation, p.4.

information, complete with examples and anonymised case studies would provide more clarity.[274]

A1.4.8 INVIVIA suggested that we should include more definitions, guidance, examples and case studies in the body of the Risk Profiles so that they are clearer. It also provided several other considerations that we should keep in mind to ensure the Risk Profiles are effective including that they should be scalable and flexible to accommodate the diversity of services, from small startups to large platforms.[275]

Our response

A1.4.9 The Risk Profiles are an input to the first step in a multi-step risk assessment process. We consider that our chosen approach, on balance, is easy for all service providers to use and can effectively incorporate new evidence as our evidence base expands. We have produced various supplementary documents and resources that should be used in tandem with the Risk Profiles. For example, the Register of Risks, the Register of Risks Glossary, Appendix A of the Risk Assessment Guidance (table of kinds of illegal content and relevant offences) and Appendix B of the Risk Assessment Guidance (examples of how to use Risk Level Tables for U2U services). We consider that these resources, taken together, will provide services with sufficient clarity, examples and guidance. As such, we do not consider it necessary to make any changes or additions to the risk profiles specifically to address these points.

Internal data

A1.4.10 5Rights Foundation stated that Ofcom should ensure that services be prepared to demonstrate that they considered their own internal data and knowledge with regards to risks on their services, in addition to the Risk Profiles.[276]

A1.4.11 The Institute for Strategic Dialogue stated that service providers should also not solely base their assessments on the Risk Profiles. Instead, it suggested they should be mandated to incorporate insights from their own internal data and understanding of user engagement, along with input from and consultation with external experts. It emphasised that the tech sector has a track record of resisting efforts to improve safety and has withheld internal evidence of risks and harms, even when superficially cooperative.[277]

Our response

A1.4.12 The role of the Risk Profiles is to help service providers understand characteristics of their service that could increase risk of harm to their users. It is not intended to specify how information is used or where the information comes from to subsequently conduct their risk assessment. We would expect for service providers to rely on a range of evidence inputs to carry out a suitable and sufficient risk assessment. This may include relevant data,

[274] Are, C response to November 2023 Illegal Harms Consultation, p.4.
[275] INVIVIA response to November 2023 Illegal Harms Consultation, pp.6-7.
[276] 5Rights Foundation response to November 2023 Illegal Harms Consultation, p.4.
[277] Institute for Strategic Dialogue to November 2023 Illegal Harms Consultation, p.7.

research and independent assessments by third parties (even if kept confidential). Table 5 in the Risk Assessment Guidance summarises the relevant types of evidence service providers should consider in Step 2 of their risk assessments. Part 3 ('Evidence') of the Risk Assessment Guidance provides more detailed information on these inputs.

Search Risk Profile

A1.4.13 Skyscanner and Mid Size Platform Group asked for clarification on what was meant by "low capacity" and "early stage" with regards to the commercial profile risk factor found in the Search Risk Profile.[278]

Our response

A1.4.14 We explain the meaning of early stage and low-capacity services in the glossary of terms found in the Register of Risks.[279]

Consideration of comparative risk assessment frameworks

A1.4.15 ACT: The App Association requested that there is maximum alignment with the Risk Profiles and ISO 31000 risk management guidelines.[280] Digital Trust & Safety Partnership provided evidence regarding its Safe Framework and its evolution which it considered would be useful to Ofcom.[281]

Our response

A1.4.16 In proposing and finalising our approach to Risk Profile and Risk Assessment Guidance, we considered various other approaches, frameworks and systems of assessing risk.

The status of the Risk Profiles under the Act

A1.4.17 The Center for Countering Digital Hate (CCDH) said that the Act treats the Risk Profiles and associated systemic issues as secondary to content judgements.[282] Some examples included were:

- "*Section 9.4 of the Act states that "as part of the assessment, services must consider various characteristics of the service specified in the legislation – such as its user base, functionalities, business model, and systems and process – and also take into account of the relevant risk profile(s) produced by Ofcom.*
- *Further, paragraph 9.73 of the November 2023 Consultation states that "Services are required to take account of our Risk Profiles when they carry out their risk assessments*

[278] Mid Size Platform Group response to November 2023 Illegal Harms Consultation, p.4; Skyscanner response to November 2023 Illegal Harms Consultation, p.11. Skyscanner response to May 2024 Consultation on Protecting Children from Harms Online, pp.10.
[279] See the Glossary of Terms in the Register of Risks.
[280] ACT: The App Association response to November 2023 Illegal Harms Consultation, p.5.
[281] Digital Trust & Safety Partnership response to November 2023 Illegal Harms Consultation, pp.10-12. Digital Trust & Safety Partnership (ofcom.org.uk)
[282] Center for Countering Digital Hate response to November 2023 Illegal Harms Consultation, p.8.

and given this, we consider their intended purpose is to help services conduct their risk assessment."

Our response

A1.4.18 The status of the Risk Profiles and the duty on services to take account of them in their risk assessment is derived from the Act. Within our legislative bounds, we consider the Risk Profiles to be a vital tool in helping service providers understand the characteristics of their service that increase the risk of certain illegal harms. We are not in a position to alter or revisit the legislative status of the Risk Profiles under the Act. That is the role of Parliament.

The connection between Risk Profiles and the Codes of Practice

A1.4.19 CCDH said there was a disconnect between the evidence of harm presented in the risk profiles and the mitigation for those harms proposed in the Codes of Practice. It explained that the Risk Profiles identified systemic, overlapping risks and offences and should therefore be a fundamental consideration rather than something to just be "taken account of". The Codes of Practice do not adequately address these aspects from the Risk Profiles and appear, to CCDH, to focus on content moderation choices and takedowns. [283]

A1.4.20 Molly Rose Foundation stated there was a disconnect between the evidence of harm presented in the Risk Profiles and Register of Risks and the mitigation for those harms proposed in the Codes of Practice.[284] It also requested a more explicit connection between the Risk Profiles, Codes of Practice and transparency metrics in the form of an annualised harm reduction framework, recommendation of appropriate measures and adopting a transparency programme largely modelled on the BEEF framework.[285]

Our response

A1.4.21 We address these and other similar points in chapter detailing 'Our approach to developing Codes measures'.

User reports

A1.4.22 Association of Police and Crime Commissioners suggested that the Risk Profiles should incorporate user complaints as they believed that they are frequently ignored by large service providers, which fuels the risk of online harms.[286]

Our response

A1.4.23 We note that user complaints are already incorporated in Step 2 of the risk assessment process of the Risk Assessment Guidance as a core evidence input, as they are likely to be

[283] Center for Countering Digital Hate response to November 2023 Consultation, p.8.
[284] Molly Rose Foundation response to November 2023 Illegal harms Consultation, p.30.
[285] Molly Rose Foundation response to November 2023 Consultation, p.44.
[286] Association of Police and Crime Commissioners response to November 2023 Illegal Harms Consultation, p.3.

specific to a particular service. As such we do not consider that it is appropriate or necessary to incorporate them as a risk factor for the purpose of our Risk Profiles.

User age and target users

A1.4.24 OnlyFans suggested that the age of users, including target users, should be a factor when considering the risks associated with a service.[287]

A1.4.25 Trust Alliance Group said that Ofcom should compel providers to state their target market and provide information to show who is actually using their service as there is a distinction between knowing who users are compared to who is using the service. This way, providers would be able to make more informed assessments of the risk presented by the operation and delivery of their service. It may be the case that there is a gap between the users that the service is targeting and the users that are actually using the service. It considered that this would be valuable information for the service and for Ofcom to have and would allow for a more accurate understanding of the risks presented by the service. If the online service provider claimed not to know who its users are or who is using the service, then this should indicate a higher risk level.[288]

Our response

A1.4.26 User age is already a consideration in the Risk Profiles under the general risk factor of user base demographics. Service providers would also be able to draw out the nuances related to their services' target users and its effects on risk of illegal harms manifesting in Step 2 of the risk assessment – where they will be able to use their own evidence.

A1.4.27 Furthermore, we cannot compel services to go beyond what is required under the Illegal Content Risk Assessment duties set out in section 9 of the Act. However, section 9(5)(f) of the Act does require service providers to take into account the different ways in which their service is used, and the impact of such use on the level of risk of harm that might be suffered by individuals.

Anonymity and pseudonymity

A1.4.28 Trustpilot asked us to make a distinction between 'anonymity' and 'pseudonymity' within the anonymous user profiles risk factor in the U2U Risk Profile as the latter does not generate the same level of risk as the former, particularly when service providers require users to create an account attached to an email address.[289]

Our response

A1.4.29 We have explained this differences between anonymity and pseudonymity in the relevant footnote of the Risk Assessment Guidance relating to risk factor 3b of the U2U Risk

[287] OnlyFans response to November 2023 Illegal Harms Consultation, p.4.
[288] Trust Alliance Group response to November 2023 Consultation, p.6.
[289] Trustpilot response to November 2023 Consultation, pp.8-9.

Profile.[290] However, we do not consider it is necessary to the explain the level of risk associated between the two in the U2U Risk Profile the place to set this out. It is for service providers to understand and explain why, for example, the ability to create anonymous user profiles is not as risky on their service through the next steps of the risk assessment process (particularly Step 2).

User verification and age assurance

A1.4.30 Trust Alliance Group also suggested that another change be made to the Risk Profiles to reflect the importance of user identity verification or assurance. It suggested user base could be included as a section within the profile to reflect Ofcom's own thinking on the subject – incorporating the child users, age assurance, anonymity and the provision of identity verification services. Alternatively, it could be incorporated into question 4 of the U2U Risk profile questions under which identity verification could be included as a crosscutting feature, impacting any type of user networking/interaction.[291]

Our response

A1.4.31 The purpose of the Risk Profiles is to help service providers understand the risk of harm connected to certain service characteristics and functionalities. Through this lens, we have accounted for the risks of children accessing a service and anonymous posting in our Risk Profiles. Different methods of verifying the age and identity of users are mitigations that providers could deploy reduce associated risks of harm.

Cultural and regional factors

A1.4.32 SWGfL proposed that the Risk Profiles should consider cultural and regional factors that influence the production and consumption of harmful content. It explained that some regions may have different attitudes towards animal cruelty, which can affect how content is created and shared, so tailoring risk mitigation strategies to specific cultural contexts can improve their effectiveness.[292]

Our response

A1.4.33 The Act does not require service providers to account for "cultural" or "regional" risk factors in their risk assessments. The matters that U2U and Search service providers are required to take account of are clearly set out in section 9 and 26 of the Act respectively.

---

[290] Footnote 57 of the Risk Assessment Guidance: We describe 'anonymous user profiles' as a user-to-user service functionality allowing users to create a user profile where their identity is unknown to an extent. This includes instances where a user's identity (an individual's formal or officially recognised identity) is unknown to other users, for example through the use of aliases ('pseudonymity'). It also includes where a user's identity may be unknown to a service, for example services that do not require users to register by creating an account. Further information on risk factors is available in the Register of Risks Glossary.

[291] Trust Alliance Group response to November 2023 Consultation, p.6. Risk Assessment Guidance, Figure 2, Question 4: Does my service have any of the following functionalities related to how users network with one another?

[292] South West Grid for Learning response to August 2024 Further Consultation, pp.4-6.

Risk of exposure

A1.4.34 SWGfL wanted greater emphasis placed on how repeated exposure to harmful content, or the presence of multiple risk factors, can lead to more significant or long-term damage in the Risk Profiles. Additionally, it suggested that the Risk Profiles could also consider indirect harms, such as how witnessing animal cruelty online might lead to future perpetration of similar acts.[293]

A1.4.35 Molly Rose Foundation suggested that, in the Risk Profiles and recommended measures, we should emphasise the importance of disrupting harm pathways where a range of design features may combine to exacerbate the risk of exposure to illegal content, including suicide and self-harm content and behaviours.[294]

A1.4.36 Molly Rose Foundation said that we fail to give due regard to how risk factors may combine or conjoin together to exacerbate risks. It was also concerned that we had inadequately reflected the cross-platform nature of harm in both the Risk Profiles and recommended measures.[295]

A1.4.37 SWGfL suggested that the Risk Profiles present the specific mechanisms by which risk factors could lead to harm in more detail. It gave the example of the social media service risk factor which could have greater detail about how algorithms that prioritise engagement might amplify the visibility of animal cruelty content and therefore increase harm. [296]

A1.4.38 SafeCast Limited was concerned that we have not considered the long-term risks in the Risk Profiles arising from bad actors that will keep and use images and records of communications made by children when they are children.[297]

Our response

A1.4.39 The Risk Profiles themselves are a tool that summarises the most important relationships between our evidence on risk factors and kinds of illegal harm. They strike a balance between being practical for service providers to use for their risk assessments and providing granular detail on how the presence of certain risk factors may lead to kinds of illegal harm to manifest on a service. We explore the link between relevant risk factors and illegal harms, including the mechanism by which the harm may manifest, specifically in the Register of Risks. This includes where the functionalities of sharing images or the availability of historic text-based communications may carry risks linked to some illegal harms in particular.

A1.4.40 We consider that the above suggestions for various ways in which risks of harm may manifest are already accounted for in our approach to the Risk Profiles. We specify several

[293] South West Grid for Learning response to August 2024 Further Consultation, pp.4-6.
[294] Molly Rose Foundation response to November 2023 Consultation, p.33.
[295] Molly Rose Foundation response to November 2023 Consultation, p.31.
[296] South West Grid for Learning response to August 2024 Further Consultation, pp.4-6.
[297] SafeCast Limited response to November 2023 Illegal Harms Consultation, p.5.

risk factors in the Risk Profiles. Hence, if a particular kind of illegal harm has multiple risk factors associated with it, we expect service providers' risk assessments to account for the heightened risk of that harm to their users if relevant. Our analysis of the causes and impacts of harms is also construed broadly, so as to include examples of evidence which demonstrate where online exposure to content might be replicated by a user in a real-world scenario. We also consider the risks connected to algorithms and repeated exposure to illegal content over time are appropriately captured by the risk factor of content recommender systems.

Concerns around only highlighting risks in the Risk Profiles

A1.4.41 Some stakeholders queried the Risk Profiles focusing only on the risks associated with characteristics, features, and functionalities of services. Broadly, they felt that, as well as highlighting risks, we should also indicate where functionalities, mitigations and other characteristics provide benefits to users and ensure that we emphasise that they are not inherently bad.[298]

Our response

A1.4.42 We believe that the Risk Profiles should continue to highlight the areas that service providers should pay close attention to in their risk assessments, which are intended to highlight the risk of harm from illegal content rather than highlight benefits of characteristics.

A1.4.43 We note that the Act itself requires the Risk Profiles to focus on risks. We also note that the benefits of certain features or functionalities and mitigations already in place can be accounted for in Step 2 of our multi-step risk assessment process. In particular, Step 2 considers how any existing controls as part of the design or operation of the service could reduce the risk of harm to users. We interpret the risk assessment duty as set out in the Act as requiring service providers to assess the actual risk of harm to users and so service providers should consider the Risk Profiles alongside actual evidence in doing so. We acknowledge the benefits in both the November 2023 Consultation and this Statement.

Continuing industry engagement

A1.4.44 TechUK recommended that we continually engage with industry to enhance understanding and interpretation of the Risk Profiles.[299]

Our response

A1.4.45 The Risk Profiles are derived from the qualitative analysis we conducted on our evidence base found in the Register of Risks. However, we envisage that through our informal

[298] Airbnb response to November 2023 Illegal Harms Consultation, p.5; Google response to November 2023 Illegal Harms Consultation, p.23-25; ICO response to November 2023 Illegal Harms Consultation, p.2. Reddit response to November 2023 Illegal Harms Consultation, p.20; techUK response to November 2023 Illegal Harms Consultation, p.14.

[299] techUK response to November 2023 Illegal Harms Consultation, p.15.

engagement with services and formal requests for information will result in us continuously improving our understanding of how we frame our Risk Profiles.

A1.5 Risk Assessment Guidance

Stakeholder responses by theme

Risk assessment requirement

A1.5.1 The Act requires all in-scope services to complete an illegal content risk assessment, and it sets out requirements which the assessment must cover.

A1.5.2 We received some feedback that expressed concern about the level of burden this duty could present:

- Safecast said that they believed that risk assessment guidance could hinder new entrants.[300]
- Protection Group International asked how risk assessments can "be undertaken solely from a UK perspective on platforms which have global reach". Further, they added that although similar risk assessments have been used in other sectors, "the online world is far more complicated" and called for additional support to help service providers comply.[301]
- Mobile Games Intelligence Forum said that the risk assessment should allow services to account for the mitigation measures in place. They also argued that "it should be possible to conduct a single risk assessment for multiple in scope services where they offer the same or similar characteristics".[302]

A1.5.3 However, we also received some feedback welcoming the introduction of risk assessments to the industry:

- Evri said that "It is a reasonable ask that organisations carry out risk assessments of products and/or services which have the potential to cause harm, or be used to cause harm, even if this is not the aim of the product or service".

Our response

A1.5.4 The requirement that service providers in scope of the Act complete an illegal content risk assessment for every service they operate is a duty set out in the Act. We have sought to write guidance to support service providers in meeting this requirement in a way which is flexible and proportionate to the range of services in scope, but where we have received responses challenging the requirement itself, we cannot adjust the scope of the duty. The purpose of the risk assessment is to assess risk as it exists on the service, and this involves considering how any aspects of the service's design or operation could affect risk –

[300] Safecast response to November 2023 Illegal Harms Consultation, p.3.
[301] Protection Group International response to November 2023 Illegal Harms Consultation, p.3.
[302] Mobile Games Intelligence Forum response to November 2023 Illegal Harms Consultation, pp.3-4.

including considering existing mitigations. The Guidance supports service providers to do this.

Proposed four-step methodology

A1.5.5 As part of our proposed Risk Assessment Guidance, we set out a four-step methodology to help services comply with relevant duties set out under the Act. The proposed Guidance is of the risk illegal harms could pose to their users.

A1.5.6 Many stakeholders are supportive of the proposals, considering it comprehensive, holistic and in line with best practice in industry.

- The Centre for Competition Policy wrote that the four-step methodology is thorough and clear.[303]
- Children's Commissioner was also supportive of the four-step methodology.[304]
- Digital Trust and Safety Partnership said that Ofcom's proposed methodology aligns with their Safe Framework assessment, including to undertake an annual risk review.[305]
- National Trading Standards E-crime Team said that "the proposed approach matches best practice and current standards in risk management, including mirroring risk assessment implemented in other sectors. For these reasons we believe it is self-evident that the proposed approach is both proportionate and appropriate".[306]
- Nexus said that the proposed "four steps outline the importance of a holistic approach to protecting users from illegal harms, from understanding the harms, assessing risk and implementing safety measures to address said risks, to reflective work and working to update assessment processes when necessary".[307]
- Segregated Payments Ltd said that the proposed "risk assessment methodology (steps and profiles) is comprehensive and it addresses the right universe of issues relating to potential harm. Applying this approach will achieve two important outcomes: correct identification of risks associated with services and, secondly, an appropriate and necessary focus on controls to minimise harmful outcomes to materialise".[308]
- Stop Scams UK said that they largely agreed with our proposals regarding risk assessments. They explained that the "four-step risk assessment process outlined provides a structured and proportionate approach for services to identify and mitigate

[303] Centre for Competition Policy response to November 2023 Illegal Harms Consultation, p.12.
[304] Children's Commissioner response to November 2023 Illegal Harms Consultation, p.21.
[305] Digital Trust and Safety Partnership response to November 2023 Illegal Harms Consultation, p.4.
[306] National Trading Standards E-crime Team response to November 2023 Illegal Harms Consultation, p.3.
[307] Nexus response to November 2023 Illegal Harms Consultation, p.5.
[308] Segregated Payments ltd response to November 2023 Illegal Harms Consultation, p.3.

online harms, including scams. By understanding the specific risks associated with their platforms, services can implement targeted safety measures to protect users".[309]

- We Protect Global Alliance said that the proposed methodology is clear and aligned with their existing framework.[310]
- ACT The App Association generally supported the proposed four-step methodology and requested maximum alignment with standardised risk management approaches.[311]
- British and Irish Law, Education and Technology Association said that the "proposed four-step risk assessment process offers a valuable framework for U2U and search services to effectively identify and manage potential online harms. However, enriching this framework with granular guidance and enhanced clarity in specific areas can further empower services to fulfil their risk mitigation responsibilities".[312]

A1.5.7 However, we also received feedback which was critical of the proposed methodology:

- An anonymous respondent said that the proposed Guidance was too prescriptive.[313]
- The Board of Deputies of British Jews agreed that the "four-steps are, prima facia, strong and set out a good mechanism for the process in which an online harm could be assessed and dealt with". However, they added that an additional step should be included which is "centred on sanctions. Either for the service that fails to adequately address the harm to the damaged party's satisfaction, or to the guilty party themselves". They explained that there could be "a huge risk here that services will not wish to adequately punish bad users and Ofcom needs to instil themselves with the power to penalise either the service or the bad actor themselves. 'Naming and shaming' is not enough. It must be in Ofcom's policies and remit to actually sanction bad users".[314]
- Trust Alliance Group expressed concern that the four-step methodology fails to take account of the significant role of safety by design. They added that consideration of design should be part of the risk assessment process.[315]
- Yoti was generally supportive of proposed methodology but stressed that Ofcom should make it clear in the Risk Assessment Guidance "that it does not recommend that providers retain the information resulting from an age assurance or age verification step taken by a user other than to assign a user to an age threshold".[316]

---

[309] Stop Scams UK response to November 2023 Illegal Harms Consultation, p.6.
[310] We Protect Global response to November 2023 Illegal Harms Consultation, p.6.
[311] ACT The App Association response to November 2023 Consultation, p.6.
[312] British and Irish Law, Education and Training Association response to November 2023 Illegal Harms Consultation, p.2.
[313] Name withheld 3 response to November 2023 Illegal Harms Consultation, p.4.
[314] The Board of Deputies of British Jews response to November 2023 Illegal Harms Consultation, p.4.
[315] Trust Alliance Group response to November 2023 Illegal Harms Consultation, p.4.
[316] Yoti response to November 2023 Illegal Harms Consultation, p.6.

Our response

A1.5.8 We welcome feedback which noted the intended flexibility of the four steps set out in the Guidance, noting the range of service providers in scope of the risk assessment duty.

A1.5.9 Regarding the Board of Deputies of British Jews' recommendation that we add a fifth step focused on sanctions, the Risk Assessment Guidance is linked closely to requirements and specific duties set out in the Act, to helps service providers meet these new duties. Adding this step would go beyond the duties set out in the Act.

A1.5.10 We have considered responses regarding the guidance being seen as too prescriptive or giving too little emphasis to safety by design. We believe the Guidance offers an appropriate level of flexibility while supporting service providers to meet the specific requirements set out in the Act. It is of course open to service providers to deviate from our Guidance, provided that they meet their risk assessment duties. We consider that safety by design is implicit in the risk assessment process we set out (such as the duty on service providers to risk assess the impact of any significant change to the design of their service before making such a change) and explicitly drawn out in some evidence inputs and guidance regarding reviewing the assessment, particularly in the event of a proposed significant change.

A1.5.11 Regarding feedback from Yoti that the Guidance should be clearer in setting expectations about retaining information and data privacy, we have pointed to relevant guidance and legislation on holding data and are clear in the Guidance that service providers are not expected to gather additional data on users to comply.

Quality of risk assessments

A1.5.12 We received some feedback expressing concern about how we will monitor or ensure the quality of risk assessments, and prevent them from becoming a tick box exercise.

- Born Free Foundation said that it is currently unclear how Ofcom will ensure that the risk assessments that service providers will be required to carry out are fit for purpose. They added that service providers should be guided specifically to consult credible and professional sources of expertise in order to adequately assess the risk of animal cruelty.[317]
- Centre for Countering Digital Hate expressed concern that Ofcom's Risk Assessment Guidance sets the wrong priorities. They explained that system design considerations should come before decisions about content and the likelihood of user encounters with illegal content. They said that this prioritisation undermines the principle of safety by design. Further, they expressed concern that the four-step methodology has the risk of being a tick-box exercise. Lastly, they doubted if Ofcom has sufficiently followed the best practice principles of risk assessment set out in Table 9.1 when explaining the

[317] Born Free Foundation response to November 2023 Illegal Harms Consultation, p.4.

four-step methodology in Table 9.2, which is less holistic and more prescriptive in their view.[318]

- Suzy Lamplugh Trust expressed concern that the regime is not outcome-oriented but instead "focused on processes that companies need to follow in a tick-box way to comply".[319]
- IICSA Changemakers said that as well as the Guidance there "remains a need to ensure that there is external input and oversight of risk assessments carried out by tech companies".[320]
- TSB Bank warned that the fact that Ofcom's methodology is not mandatory may negatively affect the quality of risk assessments and added that even those performing them in line with methodology could be seen as a 'tick box' exercise. Further, they also called for some kind of formal oversight as they believe the regime should not rely on service providers marking their own homework, and believe third party independent audits should be mandatory.[321]
- Global Partners Digital supported the Guidance but added that the severity of harm is key to orient as a priority as part of the third step. They say that this is relevant to ensure the risk assessment methodology can be tailored to match the size and context of the specific service in question.[322]
- Logically welcomed the four-step approach proposed by Ofcom. However, they urged Ofcom to reconsider not introducing third-party audits of risk assessments.[323]

Our response

A1.5.13 Regarding responses calling for the Guidance to specifically consult with experts on certain harms, such as animal cruelty, we have included consultation with experts as an enhanced input and encourage service providers to use these, if needed, to gain an accurate picture of risk on their service.

A1.5.14 On feedback which suggests the Guidance is not outcome-oriented, and concerns that the Guidance could be a tick-box exercise, we note that the Risk Assessment Guidance is intended to help providers meet their risk assessment obligations under the Act. Throughout the Guidance we have emphasised the requirement that providers rely on evidence to make their assessment and their conclusions about risk level. We have sought to follow best practice principles regarding risk assessments, while ensuring the Guidance is grounded in the requirements set out in the Act.

---

[318] Centre for Countering Digital Hate response to November 2023 Illegal Harms Consultation, p.6.
[319] Suzy Lamplugh Trust response to November 2023 Illegal Harms Consultation, p.4.
[320] IICSA Changemakers response to November 2023 Illegal Harms Consultation, p.3.
[321] TSB Bank response to November 2023 Illegal Harms Consultation, pp.2-3.
[322] Global Partners Digital response to November 2023 Illegal Harms Consultation, p.11.
[323] Logically response to November 2023 Illegal Harms Consultation, p.2.

A1.5.15 We have considered feedback expressing concern at the fact that service providers are not required to follow our guidance, and calling for third party audits to ensure standards. As set out in the Act, service providers are obliged to carry out a risk assessment and use Ofcom's Risk Profiles to do this, however, they are not required to follow Ofcom's guidance, provided they meet the requirements set out in the Act. There are not provisions in the Act to call for third party audits to ensure standards, but we will consider enforcement action regarding service providers that fail to meet their duties under the Act.

A1.5.16 Regarding the response from Global Partners Digital that severity of harm should be prioritised in the Guidance when assigning risk level and considering mitigations; we have followed the requirements set out in the Act that a service provider must consider the likelihood of illegal content occurring on the service and a user encountering it, and the nature and severity of the harm suffered.

Proportionality for larger services

A1.5.17 We received some responses which focused on the way the Risk Assessment Guidance handles large services.

- Antisemitism Policy Trust warned that a service being large should not necessarily mean the service is higher risk.

Our response

A1.5.18 The Guidance sets out that actual risk of harm to users is the key determining factor when assigning risk level. However, we also note that the size of the service's user base is a relevant consideration when determining the potential impact of a piece of illegal content, in terms of the number of users who may encounter it. We draw this out in the Guidance specifically by asking service providers to consider the impact a kind of illegal content could have by means of its service, i.e. the number of users who could encounter it.

Proportionality for smaller businesses

A1.5.19 Several respondents emphasised the need to ensure that the risk assessment is proportionate for smaller businesses.

- Federation of Small Businesses explained that they believe that risk assessments are a good way to identify and manage risks for businesses, and that the proposal that they should be carried out at least once a year or once a significant change occurs is suitable. However, they added that Risk Assessment Guidance should be flexible as "small and micro businesses are less likely than larger services to have access to more sophisticated tools to assess, manage and mitigate risks, and they should be able to reflect this in the risk assessment".[324]

- Internet Watch Foundation said that the Risk Assessment Guidance should capture "small but high-risk platforms" and ensure that the approach of the Guidance "does

[324] Federation of Small Businesses response to November 2023 Illegal Harms Consultation, p.2.

not just include “large” platforms where a lot of best practice currently exists”. Specifically, they suggest that the Guidance should consider the approach to the definition of “large platforms”, as the regulation will likely not capture some of the most popular platforms used by children and ensure that medium sized businesses are also in scope of training and development requirements for staff.[325]

- Mega warned that the volume and complexity of the consultation and Risk Assessment Guidance is counterproductive. Further, they added that it is unreasonable to impose a near identical level of obligations on smaller multi-risk services and large multi-risk services.[326]
- Canadian Centre for Child Protection warned that “child safety obligations cannot shift based on size. We do not tolerate that in the physical world, and we must keep that in mind when we assess proportionality”.[327]

Our response

A1.5.20 Regarding the feedback from the Federation of Small Businesses about the need for the Guidance to be flexible for small or microbusinesses, we think this is reflected in the Guidance sufficiently. We have considered proportionality in every step proposed in the Guidance, particularly in the sections where we provide support to service providers regarding the amount and kinds of evidence to use to support their findings and in assessing whether a proposed change could amount to a ‘significant change’ in the Act.

A1.5.21 Regarding feedback that the risk assessment should also capture small but high-risk platforms, and that we should reconsider the definition of large platforms as many popular platforms may not be captured here; we considered this feedback and reiterate that the Guidance supports providers to look to evidence of harm to users as the key determining factor when assigning a risk level - small platforms are not excluded from that. Providers are also guided to consider their user base, including whether they have vulnerable groups such as children, and the implications for risks.

A1.5.22 We considered feedback from Mega that the complexity of the consultation is counterproductive and suggesting that we have not sufficiently differentiated the obligations for small services and larger services. The requirement to complete an illegal content risk assessment is set out in the Act for all in-scope services. Where appropriate we have reflected in our Guidance some differentiation between what we consider is proportionate for smaller and larger services, so that all service providers can meet their duties in a way which is proportionate overall.

[325] Internet Watch Foundation response to November 2023 Illegal Harms Consultation, p.3.
[326] Mega response to November 2023 Illegal Harms Consultation, p.2, p.4.
[327] Canadian Centre for Child Protection response to November 2023 Illegal Harms Consultation, p.8.

Proportionality for alternative business models

A1.5.23 Some responses focused on how the risk assessment methodology could impact in-scope service providers which operate under an alternative business model.

- Wikimedia Foundation said that they perceived Ofcom's proposals as being targeted only towards "top-down, for-profit-oriented platforms and service providers". They added that, for an organisation like itself, the challenge is that it does not have the infrastructure in place to implement the processes set out in consultation. They added that the proposed requirements challenge the very nature of its self-governed platform.[328]
- Oxford Disinformation and Extremism Lab said that they are "concerned about the potential for alt-tech platforms with a significant presence of UK users and no UK-based offices or employees to evade or shirk risk assessments".[329]

Our response

A1.5.24 Regarding feedback that our proposals may pose a challenge for alternative business models to comply with, or the risk that they may not comply, we have sought to produce guidance which accommodates a range of services providers regardless of size and nature. The Guidance has been produced to help all service providers in scope meet the requirements set in the Act. Where possible we have sought to be flexible regarding how providers measure risk – such as in the range of evidence inputs they may consider to assess risk level to users. Regarding the possibility of non-compliance, we expect all service providers to comply with the illegal content risk assessment duties by the statutory deadline. Failure to do so means service providers are at risk of enforcement action by Ofcom. Information about how Ofcom will normally approach enforcement under the Act is provided in our Online Safety Enforcement Guidance.[330]

Gaps in the proposed risk assessment

A1.5.25 We received some responses which noted potential gaps in the proposed approach to the Risk Assessment Guidance.

- Glitch warned that by "not explicitly considering gender-based harm in the risk assessment process, there is a risk of overlooking the prevalence and impact of online abuse and harassment experienced by women and girls".[331]
- White Ribbon Canada called for a "stronger gender-based analysis to risk assessment and regulation of new technologies" to be applied. They recommend "a stronger,

[328] Wikimedia Foundation response to November 2023 Illegal Harms Consultation, p.11.
[329] Oxford Disinformation and Extremism Lab response to November 2023 Illegal Harms Consultation, p.3.
[330] Online Safety Enforcement Guidance, published alongside this document.
[331] Glitch response to November 2023 Illegal Harms Consultation, p.5.

proactive approach to addressing emerging threats as opposed to putting in place reactive measures".[332]

- Institute for Strategic Dialogue said that the "strong emphasis on proportionality related to potential costs incurred by services, as opposed to the impacts and costs resulting from online harms, needs to be reevaluated". They also warned that the format of the risk assessment which proposes assessing risks separately or in isolation should be amended to also urge service providers to consider how risks intersect or combine in practice.[333]
- Integrity Institute welcomed the Guidance and said that "It is useful to provide platforms with some structure to indicate what the regulator is looking for in risk assessments, and in this regard the risk profiles and four steps are a useful starting point". However, they went on to highlight that the Guidance does not propose "a 1 to 1 relationship between risk and solution, which makes it very tricky to report, so it must be clear that there is room for the platforms to discuss areas of overlapping risks and mitigations. Additionally, in this proposed process, there could be a significant lag between identifying the risk/implementing the fix and reporting it".[334]
- The Welsh Government said that they "would want to see any service targeting children and young people in Wales to be potentially considered higher risk, despite the relatively small number of users (by the definitions outlined in these proposals). This is particularly relevant when considering any services provided in the Welsh language".[335]
- [✂][336]

Our response

A1.5.26 Considering feedback from Glitch regarding gender-based harms as missing from the risk assessment, we have written the Guidance to support service providers to meet their legal duty regarding the kinds of illegal harm they need to separately assess including harms which disproportionately affect women and girls such as coercive and controlling behaviour, harassment, and intimate image abuse. There is substantial evidence about these harms in our Register of Risk, and service providers are required to consider their user base when conducting their risk assessment. Where relevant, we have guided providers to consider the impact of certain harms on vulnerable groups, including those with multiple protected characteristics. In addition, in February 2025 we will be publishing guidance for service providers looking specifically at how they can effectively address content and activity which disproportionately affects women and girls, including through their risk assessments.

---

[332] White Ribbon Canada response to November 2023 Illegal Harms Consultation, p.2.
[333] Institute for Strategic Dialogue response to November 2023 Illegal Harms Consultation, pp.6-7.
[334] Integrity Institute response to November 2023 Illegal Harms Consultation, p.6.
[335] Welsh Government response to November 2023 Illegal Harms Consultation, p.3.
[336] [✂]

A1.5.27 Regarding feedback from the Institute of Strategic Dialogue that proportionality is too focused on the impact of the risk assessment, rather than the impacts and costs of not identifying and mitigating online harms, we consider that our proportionality concerns are relevant given the nature of the duty on providers. Regarding the decision to support service providers to separately assess each kind of priority illegal content, this is the duty as set out in the Act.

A1.5.28 Regarding feedback from the Integrity Institute calling for a clear format for providers to use to complete assessments, we intentionally created flexible guidance so that providers can carry out the risk assessment in a way that works for them. On the risk of mitigations being missed as there is not one-to-one alignment between harms and mitigations, the risk assessment and safety duties are distinct duties in the Act. We have sought to align and help providers navigate both duties in the Guidance, but we must work within the framework and requirements set out in the Act.

A1.5.29 We considered the feedback from the Welsh Government, and we are aware of the risks to children and the relevance of the languages spoken by the user base and, where possible, have sought to guide service providers to consider these risks where relevant for their users.

A1.5.30 Considering feedback which asked that providers consider specific materials from experts and professionals of certain harms as part of their assessment, we consider that we have included this in the Guidance by suggesting that services should consult with experts as part of their risk assessment as an enhanced evidence input, particularly if they identify multiple specific risk factors for a certain harm.

Reviewing and updating risk assessments

A1.5.31 Some responses focused on the guidance regarding when to review or update a risk assessment.

- Samaritans expressed concern about the clarity of the Guidance and the potential burden for services. They said that the "Risk assessment reviews may be as involved and time consuming as original assessments so the wording in A5.123 around an expectation that reviews and updates 'should not be as burdensome as carrying out a new risk assessment' is unhelpful. The emphasis should be on providers ensuring that their risk assessment is reviewed at least every 12 months and is robust, irrespective of the 'burden': as the consultation document states, this is a duty that providers need to meet."[337]
- TSB Bank agreed with the proposals which said services should review their risk assessment annually.[338]

[337] Samaritans response to November 2023 Illegal Harms Consultation, p.4.
[338] TSB response to November 2023 Consultation, p. 1.

- UK Finance highlighted the need for event-driven risk assessments, they explained that illegal harm events or a new adverse MO emerging should be a trigger for risk assessment review.[339]
- [✂][340]
- Logically called for additional guidance on how service providers should update or review their risk assessment, for instance in light of changing external circumstances. They pointed to the EU Digital Services Act where "Under that law, should a crisis occur, the regulator may require large platforms to conduct a specific assessment of how their services are contributing to the crisis that has been declared".[341]
- Match Group agreed with Ofcom's proposals relating to risk assessments. However, they said that they "believe that an annual or semi-annual record reviewing process could result in unnecessary burdens for smaller developers with fewer dedicated trust and safety resources". They also expressed that "major changes to risk assessment and mitigation policies should warrant a review, rather than a more rigid approach". They added that "in-depth reviews of our records are conducted as required, but if holistic re-reviews are needed, we'd recommend recommending a period of every 2 years or longer".[342]

Our response

A1.5.32 We welcome feedback in support of the guidance regarding when to review or update a risk assessment.

A1.5.33 We considered feedback from the Samaritans and others regarding clarity of how and when service providers should review or carry out new risk assessments. The Act specifies that providers need to take steps to keep their risk assessments up to date, and when to review or carry out a new assessment in relation to a proposed change. We have sought to produce guidance which helps providers meet this duty in a way which is proportionate to their service and the potential risk to users. We consider that our Guidance is sufficiently clear that risk assessments should be robust.

A1.5.34 Regarding feedback about guiding service providers to consider event-driven risk assessments, or assessments triggered by unusual increases in certain kinds of harmful content, we have provided guidance on how providers can meet the specific duties for keeping a risk assessment up to date as set out in the Act. This is covered in Part 3 of our final Guidance, 'Making a significant change to your service'. We detail examples of the kinds of events or principles which require service providers to carry out a new risk assessment relating to a proposed change for the design or operation of the service.

---

[339] UK Finance response to November 2023 Illegal Harms Consultation, p.6.
[340] [✂]
[341] Logically response to November 2023 Consultation, p.14.
[342] Match Group response to November 2023 Illegal Harms Consultation, pp.4-5.

A1.5.35 We considered feedback that proposing service providers review their risk assessment annually could be disproportionate for smaller services. We have set out in the Guidance how a provider may go about conducting this annual review and we consider that it may not be as burdensome as completing a new risk assessment entirely, particularly if many factors have remained the same, as the feedback suggests could be the case for a smaller developer. We think that a period of 12-month review is proportionate and aligned to comparable regimes internationally or other industries.

Sharing risk assessments

A1.5.36 FCA suggested that key findings from risk assessments should be shared with relevant trusted flaggers.[343]

Our response

A1.5.37 We considered feedback that service providers should be required to share their risk assessment findings with specific groups, such as trusted flaggers. This requirement goes beyond the risk assessment duties set out in the Act, which include for some categorised services to publish share the findings of their risk assessment and to share their full risk assessment with Ofcom.

[343] FCA response to November 2023 Illegal Harms Consultation, p.5.

A1.6 Record-keeping and review

Stakeholder responses by theme

Additions to the record-keeping requirements

A1.6.1 Our Record-Keeping and Review Guidance (RK&RG) sets out the requirements of the Online Safety Act 2023 (the Act) in relation to record-keeping and our expectations about how we consider service providers can comply with these duties.

A1.6.2 Some respondents suggested additions to our guidance on written records:

- An individual respondent explained their research had found that users want access to records made about themselves, including records of decisions made on their content, and that service providers should have to provide users with access to such records.[344]
- Cybersafe Scotland recommended that the record-keeping duties should include service providers having to record the scale and process of interactions with law enforcement agencies.[345]
- Protection Group International said there should be an agreed format for record-keeping to ensure a consistent approach to records.[346]
- Safecast suggested that the largest service providers should be required to have their records independently audited, and the audit results provided to Ofcom, to prevent providers from fabricating records.[347]

Our response

A1.6.3 As a general point, our approach with the RK&RG has been to give guidance, in line with the requirements of the Act, on what we consider to be appropriate record-keeping under the relevant duties in the Act. This is because what we say is bound by what is set out in the legislation. Therefore, where we have received suggestions that go beyond what is legally required by the Act, we have generally not recommended these in the RK&RG.

A1.6.4 Regarding users' access to records of decisions made about their content, we consider that our Illegal Content Code of Practice measure concerning informing users of complaints outcomes may go some way to address this point, as complaints may include appeals

[344] Are, C response to November 2023 Illegal Harms Consultation, p.5.
[345] Cybersafe Scotland response to November 2023 Illegal Harms Consultation, p.5.
[346] Protection Group International response to November 2023 Illegal Harms Consultation, p.4.
[347] Safecast (1) response to November 2023 Illegal Harms Consultation, p.4.

about decisions taken in respect of user content.[348] This is an area we remain open to revisiting in future iterations of the Codes of Practice.

A1.6.5 Regarding the suggestion that Ofcom should tell service providers to record their interactions with law enforcement agencies, our Illegal Content Codes of Practice anticipate that some measures may involve interactions with law enforcement agencies.[349] Where service providers have implemented such measures, we expect that interactions with law enforcement agencies will be recorded, as part of the record of measures taken.

A1.6.6 Regarding the suggestion that service providers should submit their records for external audit, this is not required by the record-keeping and review duties. We note that this suggestion stems from a concern about possible fabrication of records. We have a range of tools available to us where we have concerns about the accuracy of service providers' records, including: our information gathering powers; our ability to require the appointment of a skilled person; and our powers to issue an audit notice requiring the provider to allow Ofcom to conduct an audit of its compliance and assess the risk of non-compliance.[350] Our Online Safety Information Guidance Consultation provides more detail on these powers.[351]

Person responsible for overseeing risk assessments

A1.6.7 As part of its feedback on the RK&RG one service provider, OnlyFans, requested further information on the responsibilities of the named responsible person for overseeing risk assessments, as well as their seniority.[352] [353]

Our response

A1.6.8 In brief, service providers should decide whether they want their named responsible person to be the same as for other areas of responsibility. We have not specified a level of seniority for named responsible persons, as this is likely to vary by service provider, but they should be sufficiently senior to respond to a senior governance body for compliance purposes.

---

[348] See the 'Reporting and complaints' Code measures in our Illegal Content Codes of Practice for U2U services and our Illegal Content Codes of Practice for search services.
[349] For example, measure 3E: tracking evidence of new and increasing illegal harm and measure 51: dedicated reporting channels in the Illegal Content Codes of Practice for U2U services and Illegal Content Codes of Practice for search services.
[350] Chapter 4 and Schedule 12 of the Act.
[351] Consultation: Online Safety Information Guidance.
[352] OnlyFans response to November 2023 Illegal Harms Consultation, p.3.
[353] See Chapter 3: Risk Assessment Guidance for Service Providers, which advises that service providers should have a written policy that names a person responsible for overseeing the duty to review and update risk assessments at least every 12 months.

Whether to record additional mitigation measures

A1.6.9 In response to our RK&RG, Snap asked whether, if a provider is meeting the recommendations of a relevant measure in a Code of Practice but also has additional mitigation measures in place, Ofcom will consider the additional mitigation measures as 'alternative measures'.

Our response

A1.6.10 Under the Act, a provider is to be treated as complying with a relevant duty if the provider takes or uses the measures described in a Code of Practice which are recommended for the purpose of compliance with the duty in question (to the extent that they are relevant to the provider and the service in question).[354] Having done so, additional mitigation measures that a provider may take will not constitute 'alternative measures' within the meaning of the Act, since these are actions that the provider takes when it seeks to comply with a relevant duty other than by taking or using the measures recommended in a Code of Practice for that purpose.[355]

A1.6.11 While there is no obligation on the provider to do so, in a situation where a provider implements additional mitigation measures in addition to those recommended in a Code of Practice to comply with a relevant duty, we encourage the provider to include in its record information about the additional measures in question; for example, a description of the measures, the reason for taking them, and their expected impact.

Comments out of scope of the RK&RG

A1.6.12 As part of the November 2023 Consultation responses to the RK&RG, we received comments relating to topics that are out of scope of, or not applicable to, the RK&RG and Illegal Harms Statement.

A1.6.13 5Rights Foundation suggested that providers should have to report against the harms to children and their prevalence, and that prevalence should be set by regulatory standards.[356]

A1.6.14 Google stated that it would welcome the option to self-certify that a service is likely to be accessed by children, without having to conduct a formal children's access assessment.[357]

A1.6.15 Mencap stated that they welcomed our proposals but recommended greater reference to accessible materials in the RK&RG.[358]

A1.6.16 Yoti made a series of recommendations regarding topics such as: artificial intelligence; the information retained by service providers conducting identity and verification checks; and

[354] Section 49(1) to (4) and (7) of the Act.
[355] Section 49(6) of the Act.
[356] 5Rights Foundation response to November 2023 Illegal Harms Consultation, p.17.
[357] Google response to November 2023 Illegal Harms Consultation, p.25.
[358] Mencap response to November 2023 Illegal Harms Consultation, p.5.

caveats on 'alternative measures' implemented by service providers when conducting age assessments of their users.[359]

Our response

A1.6.17 This feedback has not been addressed as part of the Illegal Harms Statement, due to it being out of scope, however, it has been considered as part of our other Consultations where relevant.

A1.6.18 Regarding 5 Rights Foundation's feedback, to the extent that responses relate to the record-keeping and review duties as they apply to children's risk assessments or safety duties, this will be addressed in our Protecting Children from Harms Online Statement, which is due to be published next year, following our May 2024 Protecting Children from Harms Online Consultation.[360]

A1.6.19 Specifically, in addition to the record-keeping and review duties, 5Rights Foundation's feedback relates to the children's risk assessment duties and proposed governance measures set out in the Protecting Children from Harms Online Consultation, which require providers to assess risk and track content harmful to children.[361]

A1.6.20 Regarding Google's suggestion, we will address comments on our proposals for children's access assessments when we publish our final Children's Access Assessments Guidance in January 2025.

A1.6.21 Regarding Mencap's feedback, the record-keeping and review duties do not require records to be published. Therefore, while we acknowledge the importance of accessibility regarding publicly available documents, we have not included further reference to accessibility in the RK&RG as our guidance applies to internal records. We will consider Mencap's feedback when producing guidance regarding information that service providers must make publicly available, such as the duty for categorised services to summarise and publish the findings of their most recent risk assessments.[362]

A1.6.22 Regarding Yoti's feedback, the topics covered go beyond the scope of the RK&RG but are relevant to our December 2023 Consultation on Guidance for Service Providers Publishing Pornographic Content, so we will consider such comments together with other feedback we have received in response to that Consultation.[363]

---

[359] Yoti response to November 2023 Illegal Harms Consultation, p.8.
[360] Consultation: Protecting Children from Harms Online.
[361] Section 11 or 28 of the Act for children's risk assessment duties and section 12 or 29 of the Act for safety duties protecting children, which our proposed governance and accountability Children's Safety Code measures derive from; in particular, proposed measures 1 and 5, as covered in Volume 4: Assessing the risks of harms to children online. The duties are requirements, while the measures are recommendations.
[362] Section 10(9) or 12(14) for Category 1 U2U services and section 27(9) or 29(9) for Category 2A search services. We intend to publish draft proposals regarding the additional duties on categorised services no later than early 2026.
[363] Consultation: Guidance for Service Providers Publishing Pornographic Content.

A1.7 Codes of Practice: Governance and accountability

Stakeholder responses by theme

Overarching feedback

A1.7.1 We received feedback on the governance and accountability measures as a package. Some responses suggested that our proposed governance measures should go further in specific areas, or should be adapted to ensure specific outcomes. Others questioned whether evidence of current best practice was appropriate, or suggested that we had not demonstrated that these measures would be effective.

- The Board of Deputies of British Jews noted that the implementation of these governance measures will only be as effective as the understanding of harm and lack of bias within organisations.[364]
- The Integrity Institute argued that our measures for governance should look at the metrics used by companies to measure safety and at how systems are tested.[365]
- Christian Action Research and Education (CARE) argued that our measures were more concerned with organisations and reputational risk than with victims.[366] CARE reiterated these comments in its response to the May 2024 Protecting Children from Harms Online Consultation ('May 2024 Consultation').[367]
- [✂][368]
- The Board of Deputies of British Jews, 5Rights Foundation, the Integrity Institute, and CARE expressed concerns that the measures needed to be strengthened and changed to ensure the effective reduction of harm on services.[369]
- 5Rights Foundation argued that our measures need to be reframed to focus on outcomes rather than measures to promote innovations in safety.[370]

[364] Board of Deputies of British Jews response to November 2023 Illegal Harms Consultation, p.2.
[365] Integrity Institute response to November 2023 Illegal Harms Consultation, pp.3-4.
[366] Christian Action Research and Education (CARE) response to November 2023 Illegal Harms Consultation, p.6; Mega response to November 2023 Illegal Harms Consultation, p.5.
[367] CARE response to May 2024 Consultation on Protecting Children from Harms Online, p.7.
[368] [✂]
[369] 5Rights Foundation response to November 2023 Illegal Harms Consultation, p.9; Board of Deputies of British Jews response to November 2023 Consultation, p.2; CARE response to November 2023 Consultation, p.6; Integrity Institute response to November 2023 Illegal Harms Consultation, p.3.
[370] 5Rights Foundation response to November 2023 Consultation, p.10.

A1.7.2 The arguments presented here are wide-ranging, but they all focus on our measures and approach not being strong or specific enough to ensure harms are properly understood and reduced across services.

Our response

A1.7.3 We have carefully considered these responses. We maintain that effective governance and accountability structures provide the foundation for service providers to identify, manage, and review online safety risks to its users. By embedding principles like accountability, oversight, independence, transparency and clarity of purpose into their operations, providers will be able to better understand and anticipate risks, increasing the likelihood that risks to users will be prioritised appropriately, and factored into strategic decision making.

A1.7.4 The governance measures should also be considered as one part of the package presented to protect people from illegal harms online – including other Codes of Practice measures, the Risk Profiles, Risk Assessment requirements and guidance, and the Illegal Content Judgements Guidance.

A1.7.5 The governance measures are in our first iteration of Codes of Practice, creating a strong foundation on which to build over time. Providers of regulated services can expect that we will continue to raise the bar over time where proportionate and appropriate to do so.

A1.7.6 The other responses referenced here make general comments about our approach to how we recommend services meet their duties under the Act. While we acknowledge that this criticism has been applied across all Codes, our approach remains the same as that which we proposed during the November 2023 Consultation. We explain the reasoning for this further in the 'Overview'.

Expansion of our measures to programmatic advertising companies

A1.7.7 The Global Disinformation Index wanted clarity as to whether programmatic advertising services would be within scope of these measures. It also highlighted that if these services are not captured there is potential risk that it does not report data that could reveal the extent to which illegal content is being funded and amplified on services.[371]

Our response

A1.7.8 The Act sets out the services that are in scope of the online safety regime. If a person provides an online service, it may be in scope of these duties. It is up to the provider to assess the nature of their service and, if necessary, seek independent specialist advice to determine whether or not their service would be subject to the requirements of the Act.

[371] Global Disinformation Index response to November 2023 Illegal Harms Consultation, p.3.

Support for smaller services

A1.7.9 Some respondents to the November 2023 Consultation and May 2024 Consultation suggested that Ofcom provide resources to support providers of small services in their duties under the Online Safety Act.

Our response

A1.7.10 We have an extensive programme of work to make the regulations accessible, and compliance more easily attainable for all providers of online services which fall in scope of the Act, which include many small or medium-sized enterprises (SMEs). We plan to launch a new 'Digital Support Service', which consists of interactive digital tools for regulated firms and are based on their perspectives and feedback. This will be accessible on the Ofcom website. Our first release will provide a four-step process for illegal harms, covering providers' risk assessment duties, Codes, and record-keeping obligations. This builds on our online safety Regulation Checker (which providers can use to check if the rules apply to their service), our 'quick guides' to the proposed regulations, and our business enquiries service. In particular, step 3 of our set of digital tools supports regulated providers in understanding which measures our Codes of Practice recommend, including governance and accountability.

Measure on annual review of risk management activities

A1.7.11 OneID argued that board reporting should include an analysis of fraud occurring on services to help improve providers' understanding of their role in mitigating it.[372]

A1.7.12 Meta noted that the measure is complementary to their commitments under the Digital Services Act ('DSA'). They stated that they would like the opportunity to comment on this further should Ofcom put forward its own review template. They advocated for a flexible proposal that affords providers with sufficient discretion in relation to the design and operation of such a review.[373]

Our response

A1.7.13 In response to OneID's point, under the Act's risk assessment requirements, providers are required to carry out an illegal content risk assessment. 'Fraud' and 'Financial Services' offences are among the risks that providers are required to risk assess. The governance board should review risk management pertaining to all illegal harm risks identified at risk assessment. Therefore, we consider that this measure, alongside the full range of measures in our Codes, will adequately address this point.

A1.7.14 In response to Meta's point, at this stage, we are not planning on recommending a review template for providers' annual reviews. We expect providers will be best placed to consider how to comply with this measure.

---

[372] OneID response to November 2023 Illegal Harms Consultation, p.1.

[373] Meta response to November 2023 Consultation, p.7.

Measures on senior accountability and on written statements of responsibilities

A1.7.15 The Association of Police and Crime Commissioners proposed the measure should specify that the accountable person should be trained in safeguarding.[374]

A1.7.16 [✂][375]

A1.7.17 Skyscanner said that the measure goes beyond what is required for senior accountability in the DSA and argued that it could impose a disproportionate burden on services in scope of both regimes.[376]

Our response

A1.7.18 In response to The Association of Police and Crime Commissioners' point, we are recommending a measure on training for individuals (other than volunteers) involved in the design and operation of a service. We consider it essential that this measure is flexible enough to be implemented by the broad range of services in scope and therefore do not think, at this stage, it would be proportionate for us to add prescriptive requirements of areas of training necessary. Each provider is best placed to develop training appropriate to their service and associated illegal harms risks.

A1.7.19 In response to [✂], at this stage, we do not have any evidence to suggest that such a measure would improve safety outcomes for users. Our Codes have been developed to make services understand and manage illegal harm risk via how their services are designed, governed and implemented.

A1.7.20 In response to Skyscanner, these are two separate regimes. We have carefully considered the impacts of this measure, and it is proportionate to expect providers to adopt it.

Measure on internal assurance and compliance functions, and on tracking evidence of new and increasing illegal harm

Measure on internal assurance

A1.7.21 [✂][377]

Our response

A1.7.22 All providers will have to complete risk assessments, which will help them to understand the harms present on their service, and identify what measures they should implement to mitigate these risks. We have included a measure in Volume 2: chapter 2: 'Content moderation' that specifies that service providers should resource their content moderation functions adequately. In doing so, we specify they should have regard to the propensity for

[374] Association of Police and Crime Commissioners response to November 2023 Illegal Harms Consultation, p.3.
[375] [✂]
[376] Skyscanner response to November 2023 Consultation, p.4.
[377] [✂]

external events to lead to a significant increase in demand for content moderation on the service.

Measure on tracking evidence of new and increasing illegal harm

A1.7.23 SPRITE+ argued that content tracking should include self-reported harms and other reporting tools, including moderation tools.[378] It noted that this would be important for immersive or virtual reality services where events are treated as communication and not content, making it difficult to evidence harms for legal action.

A1.7.24 [✂], the Financial Conduct Authority ('FCA'), Samaritans, and [✂] suggested we add a provision to this measure to ensure that any reports of new illegal content were reported externally to trusted flaggers, law enforcement, or other statutory bodies.[379] The FCA said that the findings from tracking new or unusual increases in illegal content is of substantial interest to regulatory bodies and the sharing of that information helps them respond to changing trends in fraud more effectively while also facilitating closer relationships between agencies. [✂]. [✂].

A1.7.25 C3P suggested that "in addition to the considerations of this accountability measure, the regulator should consider how investments made by larger services to improve tracking / monitoring activities can be leveraged to support smaller services. Alternatively, Ofcom should consider sourcing and / or funding solutions for smaller services to utilize".[380]

Our response

A1.7.26 In response to SPRITE+'s point, in our November 2023 Consultation, we proposed giving service providers the freedom to use a range of evidence to track illegal content, including evidence from complaints procedures and content moderation procedures. We note that providers may have to rely on external sources of information for tracking this content, as their complaints procedures may not distinguish between illegal content and other forms of content. We also note that the work on transparency reporting in the third phase of our Consultations will be relevant here in terms of accessibility to the wider public.

A1.7.27 In response to [✂], the Financial Conduct Authority ('FCA'), Samaritans, and [✂], the Act places a duty on service providers in relation to reporting CSEA content.[381] We aim to start issuing transparency notices to providers within a few weeks of the register of categorised services being finalised. Ofcom's Codes of Practice must describe measures recommended for the purposes of compliance with the duties to which they can relate.[382]

A1.7.28 In response to C3P, such a measure would require further consultation, and at this stage we do not have the evidence with which to assess the impacts.

---

[378] SPRITE+ (University of Glasgow) response to November 2023 Illegal Harms Consultation, p.3.
[379] [✂]; Financial Conduct Authority (FCA) response to November 2023 Illegal Harms Consultation, p.5; [✂]; Samaritans response to November 2023 Illegal Harms Consultation, p.4.
[380] Canadian Centre for Child Protection (C3P) response to May 2024 Consultation, p.12.
[381] Section 66 of the Act.
[382] See section 41 of the Act, in particular sub-section (10).

Measures on code of conduct, and staff compliance training

Measure on code of conduct

A1.7.29 The Canadian Centre for Child Protection (C3P) suggested that the codes of conduct should have specific inclusions that commit to child safety and the removal of child sexual abuse material (CSAM) and other illegal or harassment content.[383]

A1.7.30 Samaritans suggested we recommend that providers' codes of conduct include information on how vulnerable users posting illegal content are to be supported, with additional considerations given to supporting staff dealing with this content.[384]

A1.7.31 The Scottish Government noted that the measure did not appear to include any mechanism for external enforcement (such as from Companies House or a regulator) and wanted to better understand how the code of conduct would be monitored and reviewed.[385]

Our response

A1.7.32 In response to C3P, we note that the measure as drafted already covers the protection of users from illegal harm, and we consider that CSAM would be covered within the scope of the measure where it is applied. In terms of harassment and content which is a risk to child safety, we consider that the proposals set out in our May 2024 Consultation are a better means of mitigating these harms.

A1.7.33 In response to Samaritans, at this stage, we have not made this addition to the measure. It would require an assessment of whether including support to vulnerable users in codes of conduct can be evidenced or is likely to protect users from harm. Additionally, identifying individuals as 'vulnerable' would require the collection or creation of sensitive personal data. We will continue to assess the proportionality of potential additional measures as we plan and work towards future iterations of our Codes.

A1.7.34 In response to the Scottish Government, the Online Safety Enforcement Guidance published alongside our statement sets out how Ofcom will normally approach enforcement under the Act.

Measure on training

A1.7.35 Some respondents requested that training requirements captured specific areas:

- The Board of Deputies of British Jews and Nexus encouraged us to recommend service providers to train staff on various forms of online harms, including CSE, CSA, Safeguarding, Child Protection, and antisemitism.[386]

[383] Canadian Centre for Child Protection (C3P) response to November 2023 Consultation, p.5.
[384] Samaritans response to November 2023 Consultation, p.5.
[385] Scottish Government response to November 2023 Illegal Harms Consultation, p.3.
[386] Board of Deputies of British Jews response to November 2023 Consultation, p.2; Nexus response to November 2023 Illegal Harms Consultation, p.3.

- The Centre for Competition Policy argued that any training measures should include training on fundamental rights to ensure they are balanced within the risk management process.[387]

Our response

A1.7.36 We consider it essential that this measure is flexible enough to be implemented by the broad range of services in scope and therefore do not think, at this stage, it would be proportionate for us to add prescriptive requirements on areas training should cover. Each provider is best placed to develop training appropriate to their service and associated illegal harms risks. As a part of our measures on content moderation though, some providers are expected to provide training and materials to enable those working in content moderation to moderate content in accordance with their policies.

Additional measures not proposed

A1.7.37 In our November 2023 Consultation we sought additional evidence of the efficacy, costs and risks associated with potential future measures to i) require services to have measures to mitigate and manage illegal content risks audited by an independent third party, and ii) tie remuneration for senior managers to positive online safety outcomes.

A1.7.38 We want to thank the large number of stakeholders that have taken the time to engage with this. We have considered the responses, as summarised below. At this time, we have decided not to add any additional measures to what we consulted on in November 2023. We will continue to assess the proportionality of potential measures as we plan and work towards future iterations of our Codes.

Third-party auditing

- Name Withheld – a civil society organisation disagreed with our decision not to propose that providers use third-party auditing.
- EticasAI disagreed with us not including this proposal, stating that it is the most effective step towards enforcement and implementation. It noted this has been proven successful in various industries.[388]
- The Independent Inquiry into Child Sexual Abuse (IICSA) Changemakers stated that external input and oversight of risk assessments is important given the risk that some providers could seek to bury or underplay evidence of harm.[389]
- The Institute for Strategic Dialogue argued that third-party audits should be mandated from the outset given the slow pace of change towards safety and the evasive behaviours of several providers.[390]

---

[387] Centre for Competition Policy response to November 2023 Illegal Harms Consultation, p.5.
[388] EticasAI response to November 2023 Illegal Harms Consultation, p.2.
[389] Independent Inquiry into Child Sexual Abuse (IICSA) Changemakers response to November 2023 Illegal Harms Consultation, p.3.
[390] The Institute for Strategic Dialogue response to November Illegal Harms 2023 Consultation, p.6.

- Trust Alliance Group welcomed that independent third-party auditing is an option that providers could take to ensure that measures taken to mitigate illegal harms are effective.[391]
- One stakeholder emphasised that audits must be independent and tailored to the specific risks associated with each online service. It shared that it currently instructs an independent third party to assess and validate the design, implementation, and effectiveness of the compliance program. It also shared that this comes at a substantial financial cost and requires significant time from senior leaders in the business.[392]
- The NSPCC urged Ofcom to reconsider the involvement of independent third-parties in monitoring and assurance for large multi-risk services. It supported the function, stating that external review can help ensure that risk management systems fully meet the requirements of the regulation, identify where companies need to strengthen processes, and bolster Ofcom's supervision efforts. It also stated that external assurance can help improved transparency in regulatory regimes, referencing the water sector as an example. It shared the view that third-party auditing will be particularly valuable for services choosing to implement their own measures, rather than comply with the Codes of Practice, due to challenges involved in evaluating whether the measures are reasonable equivalents of those recommended in the Codes. It shared the view that it would be reasonable to expect that those in scope have the resource to fund a third-party review as this recommendation is for large multi-risk services. And given their mitigation processes will likely be the most complex (due to size and risk level), they also stand to benefit most from an external evaluation.[393]
- One respondent raised concerns with requiring companies to have their illegal content risks audited by independent third parties. They argued that whilst this would theoretically hold companies accountable and increase the transparency and reliability of reports, this approach could come with significant drawbacks. For example, independent auditing could become outdated rapidly, whereas allowing companies to identify new trends in online harms retains knowledge and skills and results in a quicker implementation of solutions.[394]
- ACT The App Association agreed with Ofcom's decision not yet to make recommendations on this due to limitations in currently available evidence.[395]
- CELE recommended that were Ofcom to consider such a measure, Ofcom should adopt the requirements set out in Article 37 of the Digital Services Act. It also suggested that Ofcom should require auditors to have proven experience in human rights impact assessments and provide guidance on how audits can comply with the Act's goals.[396]

---

[391] Trust Alliance Group response to November 2023 Illegal Harms Consultation, p.3.
[392] OnlyFans response to November 2023 Illegal Harms Consultation, p.3.
[393] NSPCC response to November 2023 Illegal Harms Consultation, p.7-8.
[394] Match Group response to November 2023 Illegal Harms Consultation, p.3.
[395] ACT The App Association response to November 2023 Illegal Harms Consultation, p.5. We emphasise this point again in paragraph 5.184 in the 'Governance and Accountability' chapter in Volume 1.
[396] CELE response to November 2023 Illegal Harms Consultation, p.4.

- [✂][397]

Linking renumeration to online safety outcomes

- NSPCC encouraged Ofcom to consider how they can shift senior management attention to safety outcomes, stating that in other sectors executive remuneration is tied to delivering positive outcomes.[398] It noted though that in these sectors the regulator set the revenue.
- LinkedIn suggested that if Ofcom were to propose such a measure, it should focus on company-wide responsibility to better ensure the right level of corporate incentives and help avoid unintended behaviours and outcomes.[399]
- Name Withheld 3 highlighted that some providers of large services would have very complex lines of responsibility as they are made up of multiple different companies providing services in a mesh.[400] This would make it unclear Ofcom would be targeting or why.
- [✂]
- Microsoft recommended against such a measure as a significant determination of remuneration, particularly if they are rate-based outcomes, though encouraged a focus on tying it to the demonstration of positive online safety behaviours were Ofcom to consider it in the future.[401] Microsoft stated that rate-based safety outcomes could include things out of a senior manager's control, could make determining compliance difficult, and could even discourage managers from documenting online harm risks.
- Name Withheld 4 stated that 'positive online safety outcomes' is an undefined term which would be difficult to accurately measure in practice, and that it was not aware of research-backed evidence supporting such a measure.[402]
- Protection Group International queried who would take overall relevant responsibility when within platforms there is a 'struggle of power between Product, Policy, Operations and Legal teams'.[403]
- Safe Space One queried whether any serious investigation of the safety critical systems market had been reviewed in this regard.[404] It stated that there are measurable objectives for safety in these sectors and suggested that there may be additional motivations beyond measurement versus industry accepted safety standards there.
- CELE agreed with Ofcom's decision not to propose this measure at this point, suggesting that it could incentivise the over-moderation of borderline and permitted content or inaccurate reporting of online safety matters by regulated companies.[405]

---

[397] [✂].
[398] NSPCC response to November 2023 Illegal Harms Consultation, p.8.
[399] LinkedIn response to November 2023 Illegal Harms Consultation, p.5.
[400] Name Withheld 3 response to November 2023 Illegal Harms Consultation, p.4.
[401] Microsoft response to November 2023 Illegal Harms Consultation, pp.5-6.
[402] Name Withheld 4 response to November 2023 Illegal Harms Consultation, p.2.
[403] Protection Group International response to November 2023 Illegal Harms Consultation, p.3.
[404] Safe Space One response to November 2023 Illegal Harms Consultation, pp.4-5.
[405] Centro de Estudios en Libertad de Expresion y Acceso a la Informacion (CELE) response to November 2023 Illegal Harms Consultation, pp.4-5.

- The Centre for Countering Digital Hate (CCDH) said that Ofcom was right to identify the misalignment of business incentives and online safety, but that the primary question should be what accountability such a measure would provide to rectify this problem.[406]
- Google recognised the importance of structural incentives, but said that such a measure would be neither workable in practice nor consistent with the scheme of the Act.[407] It stated that such a measure would be difficult to compare consistently and reliably between services and that the Act was premised on services having different risk levels with residual risk profiles varying by the inherent attributes of services. It argued that such a measure could incentivise the over-removal of content, and encourage autocratic regimes to draft legislation in a way that undermined human rights. It added that the impact of regulation on individual staff and talent acquisition and retention is an important factor in location decisions, noting the UK Government's desire to develop the UK into a 'tech superpower'.
- ACT The App Association agreed with Ofcom's decision not yet to make recommendations on this due to limitations in currently available evidence.[408] It also questioned whether this would be an appropriate way of achieving the UK Government's goals in creating the Act and what precedent it could set.

Our response

A1.7.1 These responses raise a number of questions about the practicability of linking remuneration to safety outcomes. Bearing this in mind and given the responses did not present clear evidence of the benefits of linking remuneration to safety outcomes we have decided not to pursue the idea of including a measure on this in our Codes at this time.

[406] Centre for Countering Digital Hate (CCDH) response to November 2023 Illegal Harms Consultation, p.5.
[407] Google response to November 2023 Illegal Harms Consultation, pp.12-13.
[408] ACT The App Association response to November 2023 Illegal Harms Consultation, pp.4-5.

A1.8 Codes of Practice: Content moderation

Stakeholder responses by theme

Cross-cutting themes

Additional measures proposed

A1.8.1 A number of stakeholders proposed a range of different measures to be added to the Codes of Practice ('Codes').

- The National Society for the Prevention of Cruelty to Children (NSPCC) suggested that online services should be required to quality-assure their moderation systems and adjust the balance between human and automated moderation to ensure correct outcomes.[409]
- The Scottish Government requested additional guidance to encourage providers to publish content moderation standards and performance to give users choice on whether to engage with a service.[410]
- Some stakeholders proposed additional mechanisms to record decisions about content removal. Christchurch Call Advisory Network suggested an evidence preservation mechanism for removed content.[411] An individual expressed the need for users whose content has been taken down to be able to access their case records and a record of decisions made on their content.[412]
- UK Finance emphasised the need for human moderators to proactively engage based on customer complaints or trusted flaggers, as automated systems can fail to recognise criminal intent in reported posts.[413]
- Yoti proposed a requirement for service providers to verify that all new actors uploading adult content are over 18 and consensual.[414]
- Glitch highlighted a lack of gender-specific content moderation measures in our Codes, and that this fails to recognise and address gender-based risks.[415]
- Logically requested detailed guidance on applying automated and manual content moderation to mitigate the risk of foreign interference offences.[416]

[409] NSPCC response to November 2023 Consultation, p.20. We note that NSPCC made a similar point in its response to the May 2024 Consultation on Protecting Children from Harms Online, p.48.
[410] Scottish Government response to November 2023 Illegal Harms Consultation, p.6.
[411] Christchurch Call Advisory Network response to November 2023 Illegal Harms Consultation, p.3.
[412] Are, C. response to November 2023 Consultation, p.5.
[413] UK Finance response to November 2023 Illegal Harms Consultation, p.8.
[414] Yoti response to November 2023 Illegal Harms Consultation, p.14.
[415] Glitch response to November 2023 Illegal Harms Consultation, p.6.
[416] Logically response to November 2023 Consultation, p.17.

- A number of stakeholders recommended that service providers signpost support to users whose content has been removed, particularly for suicide or self-harm content.[417]
- The NSPCC suggested that we should recommend a range of tools, including human moderation, to both prevent and disrupt grooming.[418]
- The Independent Reviewer of Terrorism Legislation recommended that service providers should prioritise content moderation practices that prevent children from encountering terrorism content.[419]
- An individual argued that training should be extended to users in order to understand the violations they have committed.[420]
- [✂][421]
- Some stakeholders recommended that service providers consult experts on illegal harms to assist with moderation processes.[422]
- In response to our May 2024 Consultation, the Northern Ireland Commissioner for Children and Young People (NICCY) said it would be helpful to see an annual moderation report from companies subject to the new Codes.[423]

Our response

A1.8.2 We recognise that a number of stakeholders asked for additional measures in this area. As noted in the chapter 'Our approach to developing Codes measures', our strategy is to recommend measures quickly in order to not delay protections for users. To recommend additional measures would require us to carry out additional impact assessments and hold a further consultation. Therefore, at this time we have decided not to add any additional measures to those we consulted on in November 2023. We will continue to assess the proportionality of potential measures as we plan and work towards future iterations of our Codes.

Self-governing communities

A1.8.3 Stakeholder feedback highlighted the importance of ensuring that our proposals take into account public interest platforms. The Wikimedia Foundation emphasised the need for our proposals to consider services created and curated by self-governing communities (such as those hosting free educational content) which are usually based on publicly available policies that these communities enforce themselves.[424]

---

[417] Graham, Dr R. response to November 2023 Illegal Harms Consultation, p.2; NSPCC response to November 2023 Consultation, p.51; Samaritans response to November 2023 Illegal Harms Consultation, p.6; Scottish Government response to November 2023 Consultation, p.6.
[418] NSPCC response to November 2023 Consultation, p.27.
[419] The Independent Reviewer of Terrorism Legislation response to November 2023 Illegal Harms Consultation, p.7.
[420] Are, C. response to the November 2023 Consultation, p.7.
[421] [✂]
[422] Born Free Foundation response to November 2023 Consultation, p.5; Four Paws response to November 2023 Illegal Harms Consultation, p.13.
[423] NICCY response to May 2024 Consultation on Protecting Children from Harms Online, p.33
[424] Wikimedia Foundation response to November 2023 Consultation, p.21.

Our response

A1.8.4 While we note this argument, we consider our measures to be technically feasible and proportionate for self-governing services. We are therefore not excluding providers of self-governing services like Wikimedia Foundation from these measures.

General monitoring

A1.8.5 Several stakeholders requested clarification that the Codes would not impose a 'general monitoring' obligation, which would require service providers to proactively monitor for illegal content.[425] Google emphasised that this clarity was needed to reduce the risk of over-removal of legal content.[426] Roblox noted that general monitoring would interfere with intermediary liability exemptions available in various legal regimes across the world, which exempt intermediaries from being held legally responsible for user activity to facilitate smooth functioning of the internet.[427]

Our response

A1.8.6 We do not have the powers to determine if measures constitute a general monitoring obligation. Our role is limited to making recommendations on how services should comply with their duties under the Online Safety Act 2023 ('the Act').

Proactive detection of content

A1.8.7 We received opposing stakeholder views about whether our measures should recommend proactive detection of content.

A1.8.8 Several stakeholders raised concerns about requiring service providers to proactively review content. Safe Space One highlighted the negative privacy implications of reviewing content before it is uploaded.[428] Evri requested clarity on whether content moderation involves checking content prior to it being made available, noting the significant burden associated with it.[429] Big Brother Watch and Global Partners Digital supported applying measures only to illegal content that providers are aware of.[430]

A1.8.9 In contrast, some stakeholders were concerned about providers' reliance on user reporting to identify illegal content and supported proactive content detection.[431] The Marie Collins

---

[425] Airbnb response to November 2023 Consultation, p.14; Google response to November 2023 Consultation, pp.33-34; Roblox response to November 2023 Illegal Harms Consultation, p.14.
[426] Google response to November 2023 Consultation, pp.33-34. We note that Google made a similar point in its response to the May 2024 Consultation on Protecting Children from Harms Online, p.26.
[427] Roblox response to November 2023 Illegal Harms Consultation, p.14.
[428] Safe Space One response to November 2023 Illegal Harms Consultation, p.11.
[429] Evri response to November 2023 Consultation, p.5.
[430] Big Brother Watch response to November 2023 Consultation, p.2. We note that Big Brother Watch made a similar point in its response to the May 2024 Protecting Children from Harms Online Consultation, p.35; Global Partners Digital response to November 2023 Consultation, p.13.
[431] Institute for Strategic Dialogue response to November 2023 Consultation, p.10; Marie Collins Foundation response to November 2023 Illegal Harms Consultation, pp.9-10; Online Safety Act Network (OSA Network) Annex B response to November 2023 Illegal Harms Consultation, pp.13-14. We note that the Commissioner

Foundation suggested pre-screening content to prevent child sexual abuse material (CSAM).[432] The Molly Rose Foundation highlighted the need for measures to include proactively detecting certain forms of priority illegal content, such as suicide and self-harm content.[433] Nexus emphasised that the effectiveness of our proposals depends on providers' ability to detect illegal content and users' ability to report it.[434] [✂][435]

Our response

A1.8.10 In chapter 4 of Volume 2: 'Automated content moderation' we recommend that some providers use specific automated content moderation methods to detect specific types of harm (for example, we recommend that some providers use hash-matching technology for the detection of CSAM).

A1.8.11 We note stakeholder arguments that we could go further in our recommendations on the proactive detection of illegal content for moderation. However, at this stage we are not in a position to go further. As set out in 'Our approach to developing Codes measures' our strategy is to recommend measures quickly in order to not delay protections for users. To recommend additional measures would require us to carry out additional impact assessments and hold a further consultation. Therefore, at this time, we have decided not to add any additional measures in addition to those we consulted on in November 2023, including on the proactive detection of content. However, we are currently considering evidence surrounding the use of automated tools to proactively detect illegal content and the content most harmful to children, going beyond the automated detection measures we have already consulted on. We intend on consulting on these additional measures in Spring 2025.

Safeguarding and pay of moderators

A1.8.12 Several stakeholders expressed concerns about the wellbeing impacts of the Codes on content moderators and suggested additional safeguarding requirements.[436] Zevo Health, Protection Group International, and Big Brother Watch raised concerns that performance targets could negatively impact moderator wellbeing.[437] Big Brother Watch cited evidence of existing workplace stress, trauma, and pressure on moderators, arguing that increased

---

Designate for Victims of Crime Northern Ireland made a similar point in its response to the May 2024 Consultation on Protecting Children from Harms Online, p.5

[432] Marie Collins Foundation response to November 2023 Consultation, pp.9-10.

[433] Molly Rose Foundation response to November 2023 Illegal Harms Consultation, p.35.

[434] Nexus response to November 2023 Illegal Harms Consultation, p.9.

[435] [✂]. [✂].

[436] Big Brother Watch response to November 2023 Consultation, pp.4-5; British and Irish Law Education Technology Association (BILETA) response to November 2023 Illegal Harms Consultation, pp.9-10; Name withheld 5 response to November 2023 Illegal Harms Consultation, p.9, Global Partners Digital response to November 2023 Illegal Harms Consultation, p.12, Protection Group International response to November 2023 Illegal Harms Consultation, p.6, SPRITE+ (York St John University) response to November 2023 Illegal Harms Consultation, p.9; Zevo Health response to November 2023 Illegal Harms Consultation, p.8.

[437] Big Brother Watch response to November 2023 Consultation, pp.4-5; Protection Group International response to November 2023 Consultation, p.6; Zevo Health response to November 2023 Consultation, p.8.

requirements for moderation exacerbate these issues.[438] Global Partners Digital recommended additional requirements on pay, work quotas, and wellbeing support for both in-house and outsourced moderators.[439] The British and Irish Law Education Technology Association (BILETA) and [✂] highlighted the importance of moderator training and wellbeing for the success of content moderation functions (but did not provide any direct evidence of this link).[440]

Our response

A1.8.13 It is not within our powers to make recommendations about moderator wellbeing unless there is evidence of an effect on user safety. Stakeholders did not provide evidence of such an effect in their responses.

Measures should be more preventative

A1.8.14 Several stakeholders highlighted that content moderation is primarily a reactive measure and over-reliance on taking down content can have limitations to protecting users online. The Institute for Strategic Dialogue argued that the variety and prevalence of online harms make content moderation necessary but not always sufficient in mitigating risks. It said that content moderation is a primarily reactive measure which should not be over-relied upon over preventative safety-by-design efforts.[441] 5Rights Foundation emphasised that content moderation should not be prioritised over upstream preventive measures.[442] End Violence Against Women Coalition called for better understanding of the direct harms to women and girls, rather than focusing solely on content takedown.[443] Yoti suggested that we should focus on the design and operation of systems, not just on content removal.[444] Christian Action Research and Education (CARE) highlighted that content moderation is focused on take-down, rather than prevention and that there is no obligation to ensure that providers design their services in a way that mitigates risk.[445]

Our response

A1.8.15 Overall, our measures cover a range of aspects of the design and operation of an online service. We recognise effective content moderation as one aspect of this and we see it as a necessary part of a range of measures that providers should implement to ensure users remain safe.

---

[438] Big Brother Watch response to November 2023 Consultation, pp.4-5.
[439] Global Partners Digital response to November 2023 Consultation, p.13.
[440] BILETA response to November 2023 Consultation, pp.9-10; Name withheld 5 response to November 2023 Consultation, p.9.
[441] Institute for Strategic Dialogue response to November 2023 Consultation, p.9.
[442] 5Rights Foundation response to November 2023 Illegal Harms Consultation, p.21.
[443] End Violence Against Women Coalition response to November 2023 Illegal Harms Consultation, p.3.
[444] Yoti response to November 2023 Consultation, p.13.
[445] Christian Action Research and Education (CARE) response to November 2023 Illegal Harms Consultation, p.9.

Harmful content

A1.8.16 We received stakeholder feedback asking for clarity on what constitutes harmful content. SPRITE+ requested more detail on the categorisation of harmful content and where the boundaries are between this and unharmful content.[446]

Our response

A1.8.17 The Act outlines a number of illegal harms which are discussed in detail within our Register of Risks ('Register') and the illegal content judgements guidance. Therefore, we are not planning to add more detail on the definition of "harmful content" within the Codes.

Small services and the risk of illegal content

A1.8.18 We received feedback noting that very small, low risk services host lower amounts of illegal content than large, multi-risk services. One individual stated that, in their personal experience, very small low-risk services do not host illegal content.[447]

Our response

A1.8.19 This point has been taken into account in our decision to apply the most onerous measures in the Codes to large and multi-risk service providers only. However, we consider that all services should have systems and processes in place to review and (if it is technically possible to do so) take down illegal content of which they are aware, in order to meet the requirements of the Act. We outline our reasoning for this in more detail in chapter 2 of Volume 2: 'Content moderation' in the section entitled 'Who this measure applies to' under the measures on reviewing, assessing and swiftly taking down content.

Overreach of the measures on legitimate actors

A1.8.20 We received feedback about the potential risk of existing proposals extending to legitimate actors. The Oxford Disinformation and Extremism Lab raised concerns about the scope of the proposals potentially capturing legitimate peaceful dissent, civil society, and academic researchers. It urged for more targeted scoping and clearer definitions to prevent this.[448]

Our response

A1.8.21 It is a requirement of the Act that providers have in place proportionate content moderation systems and processes to swiftly take down illegal content. In chapter 2 of Volume 2: 'Content moderation' (paragraph 2.220) we note that the performance targets measure does not encourage considering nuanced content in less detail, and the same would apply to content produced by legitimate actors. There are safeguards in this measure which mitigate against this risk, such as the requirement to balance the need of moderating content swiftly with the importance of making accurate moderation decisions. However, this is ultimately a risk that we are obliged to take given our duties under the Act.

---

[446] SPRITE+ (York St John University) response to November 2023 Illegal Harms Consultation, p.9.
[447] Bolton, C. response to November 2023 Illegal Harms Consultation, p.5.
[448] Oxford Disinformation and Extremism Lab response to November 2023 Illegal Harms Consultation, p.8.

How providers comply with the measures

A1.8.22 We received feedback on how providers should comply with the content moderation measures. The New Zealand Classification Office highlighted the importance of providers grounding their compliance processes in evidence, and with reference to fair, evaluative criteria to determine the risk of harm to individuals and communities.[449]

Our response

A1.8.23 We note this point and consider that the approach we have taken to the Codes of setting out (where relevant) the factors to which providers should have regard when designing their content moderation systems and processes goes some way in addressing this. For example, our measures on internal content policies, prioritisation, resourcing, training of individuals working in moderation, and the provision of materials to volunteer moderators all recommend that providers have regard to their risk assessments to be in accordance with these measures. We consider that a provider's risk assessment is grounded in evidence and allows providers to determine the risk of harm to individuals and communities on their particular services.

Review and updating of the content moderation measures

A1.8.24 We received stakeholder feedback highlighting the importance of reviewing and updating the content moderation Codes. INVIVIA suggested that the definition and categorisation criteria should be subject to regular review and adaptation, so that it remains effective in tackling rapidly evolving online threats.[450] In response to the May 2024 Consultation of Protecting Children from Harms Online ('May 2024 Consultation'), the Mid Size Platform Group highlighted that regulation is constantly evolving, and that we should consider how this content moderation can be regulated in a stable manner, consulting with industry in putting together a long term plan around this.[451]

Our response

A1.8.25 We agree with the importance of ensuring the Codes are robust and effective. In 'Our approach to developing Codes measures' we state that we plan to build on the Codes and have already committed to a further consultation in Spring 2025. However, we do not think every new threat should trigger a review of our Codes. It is for providers to identify new threats and review how they implement these Codes measures accordingly.

Content providers and infrastructure providers

A1.8.26 Global Network Initiative argued that the Online Safety Act does not cover infrastructure providers but highlighted that content they host is part of the definition of harmful content. It emphasised the importance of recognising that different kinds of services have

[449] New Zealand Classification Office response to November 2023 Consultation, p.7.
[450] INVIVIA response to November 2023 Illegal Harms Consultation, p.13.
[451] Mid Size Platform Group response to May 2024 Consultation on Protecting Children from Harms Online, p.3

different abilities to control or moderate content on their platforms, due to the distinction between direct content providers and infrastructure providers.[452]

Our response

A1.8.27 As set out in 'Our approach to developing codes measures', our measures apply to the 'provider' of the service as defined in the Act, regardless of the type of service. Where more than one person has a role in relation to the provision of a regulated service, there may be uncertainty which one is the 'provider' of the service. It is for the persons concerned to manage this risk, where appropriate by taking legal advice, in order to ensure that the duties are met.

Measures on reviewing, assessing and swiftly taking down content

Relationship between terms of service and content moderation

A1.8.28 Some stakeholders requested clarity on whether service providers have the discretion to remove content that only violates their terms of service, rather than priority illegal content. 5Rights Foundation noted that the legal requirement to remove illegal content overrides service providers' discretion to remove content based solely on their terms and conditions.[453] The Association of British Insurers suggested that we should mandate the removal of illegal content rather than leaving it to the discretion of the individual service, as current inconsistencies in handling false information, hate speech, and other illegal harms create opportunities for perpetrators of fraud.[454]

A1.8.29 In response to an equivalent measure proposed in the May 2024 Consultation, Centre to End All Sexual Exploitation (CEASE) expressed concerns about our recommendation that providers should enforce their own terms of service in moderating content, highlighting evidence that providers do not do this currently.[455]

Our response

A1.8.30 We consider that 5Rights Foundation and the Association of British Insurers have misinterpreted the meaning of the option to assess content through providers' terms of service in the measures on reviewing, assessing and swiftly taking down content (as outlined in chapter 2, Volume 2: 'Content moderation'). As set out in paragraph 2.48, providers can choose to remove content if it violates their terms of service, but this only applies when the provider is sure that its terms prohibit the types of illegal content defined in the Act.

A1.8.31 We consider that the concerns expressed by CEASE about the enforcement of terms of service by providers are addressed by the fact that the Act requires that terms of service

[452] Global Network Initiative response to November 2023 Consultation, p.8
[453] 5Rights Foundation response to November 2023 Consultation, p.21.
[454] Association of British Insurers response to November 2023 Illegal Harms Consultation, p.4.
[455] CEASE response to May 2024 Consultation on Protecting Children from Harms Online, pp.17-18

are consistently applied so that all users will understand how they will be protected from illegal content.[456] Our Codes are a package of measures that we recommend providers implement to comply with their duties.

Requirement about levels of moderation within the measure

A1.8.32 [✂] highlighted that this measure did not include [✂].[457]

Our response

A1.8.33 Given the large variety of services in scope of this measure, and the significant risk of unintended consequences if we seek to be prescriptive at this stage in the establishment of the regulatory regime, we do not consider that it would be proportionate for us to be prescriptive about the levels of moderation dependent on a provider's user base. Two services with a similar number of users or similar userbase characteristics may have different functionalities and risks, and therefore require different levels of moderation.

A1.8.34 We recommend that providers of multi-risk and large services resource their content moderation function so as to give effect to their internal content policies and performance targets. We also recommend that providers have regard to the needs of their UK user base in relation to languages. We consider that these recommendations will incentivise providers to have the appropriate level of moderation to protect users on their services, taking into account a more diverse array of factors than just the number of users on their service or the characteristics of their user base.

Measure on internal content policies

Emerging harm

A1.8.35 A few stakeholders commented on the consideration of emerging harm in internal content policies. Google suggested that emerging harm should be part of the risk assessment rather than a standalone factor in policy development, due to the challenges in assessing how specific harms manifest. It provided an example of YouTube's focus on tracking emerging trends in inappropriate content and preparing teams to address them early.[458] The Online Safety Act Network (OSA Network) also highlighted "signals of emerging harm" as an issue, providing evidence of inadequate content policies in large companies.[459]

Our response

A1.8.36 We have addressed these concerns through an amendment to the measure on internal content policies (consulted on as an addendum to the May 2024 Consultation) which clarifies that the recommendation to track emerging harm refers to an existing recommendation (outlined in Volume 1: chapter 5: 'Governance and accountability') that providers track evidence of new and increasing illegal harm on the service, separate to

[456] Sections 10 (2), 10 (3), 10 (7) and 21 (3) for U2U services of the Online Safety Act 2023.
[457] [✂].
[458] Google response to November 2023 Consultation, p.34.
[459] OSA Network Annex B response to November 2023 Consultation, p.12.

their risk assessments. It is reasonable to infer that providers would need to have processes in place for updating their policies in response to the data they are already collecting on evidence of new and increasing illegal harms.

Safety by design

A1.8.37 We received feedback on the need for service providers to prioritise safety by design in their internal content policies. The Molly Rose Foundation suggested that providers should operationalise their policies to control how quickly content spreads, including how and under what circumstances material may be de-ranked, down-weighted or deemed unsuitable for algorithmic recommendation.[460]

Our response

A1.8.38 In developing the Codes, we identified recommender systems[461] as a risk factor for some illegal harms. We also acknowledge that there are significant overlaps between those harms and primary priority content for children. Therefore, we consider that the measure on recommender systems proposed in the May 2024 Consultation may cover some aspects of relevant illegal content.

A1.8.39 However, we have since published an open letter on online services in relation to the violent riots that took place following the tragic murders in Southport.[462] Posts about the Southport incident and subsequent events from high-profile accounts reached millions of users, demonstrating the role that virality and algorithmic recommendations can play in driving divisive narratives in a crisis period. We are therefore considering the evidence base for potential new measures on recommender systems in relation to illegal harms.

Effectiveness of internal content policies

A1.8.40 Protection Group International noted that while many service providers already have internal policies and processes in place, they are ineffective due to a lack of alignment and coordination in content moderation processes.[463]

Our response

A1.8.41 Until the safety duty comes into force around March 2025, providers are not under an obligation to have effective policies. The Act itself could help to address Protection Group International's concerns. To the extent that the concern is that providers are not applying their internal content policies to all content, this is likely to be a flaw in another aspect of their content moderation systems and processes. Our recommended measures on the training of individuals working in moderation and performance targets may help to address this concern, as well as other aspects of the Act such as Section 72(2): the duty on Category

[460] Molly Rose Foundation response to November 2023 Consultation, p.35.
[461] As explained in chapter 7 of Volume 2: 'Recommender systems', recommender systems are a primary means through which user-generated content is disseminated across U2U services, and the means via which users encounter content.
[462] Ofcom, 2024. Letter from Dame Melanie Dawes to the Secretary of State, 22 October 2024. [accessed 27 November 2024]
[463] Protection Group International response to November 2023 Consultation, p.6.

1 services to apply their terms and conditions. As this is the beginning of the regulatory regime, we will revisit and improve on our Codes (including on the concerns raised by the Protection Group International) over time.

Measure on performance targets

Comparing providers' performance with performance targets

A1.8.42 We received opposing stakeholder views on setting performance targets for content moderation.

A1.8.43 Some stakeholders noted the risks associated with performance targets. Reddit and the Mid Size Platform Group expressed concerns that performance targets could be reductive, failing to account for different business sizes and models. They also highlighted that such targets might incentivise over-moderation, leading to unnecessary takedowns, threats to free expression, and disproportionate burdens on smaller companies.[464]

A1.8.44 In contrast, other stakeholders supported the publication and comparison of different providers' performance targets. Refuge recommended we specify targets for different platforms and include robust processes for transparency reporting to ensure such targets are reviewed, monitored, and published.[465] The Institute for Strategic Dialogue noted that many existing transparency reports do not present an objective assessment as they rely on self-selected metrics, and recommended that we set consistent performance metrics to allow for cross-industry comparison.[466]

Our response

A1.8.45 As outlined in paragraphs 2.213 and 2.218 in Volume 2: chapter 2: 'Content moderation', we give providers the flexibility to set different metrics in their performance targets, different targets for different types of content, and set targets at a level appropriate for their service. We consider that this flexibility means that for the time being it is likely that performance targets will vary across services.

A1.8.46 Our consultation on our draft statutory transparency reporting guidance, published in July 2024, covers the process that we will adopt for deciding what providers must include in their transparency reports. As outlined in Volume 2: chapter 2: 'Content moderation' (paragraph 2.204), we have not yet assessed the evidence to determine if it would be feasible and beneficial for us to set parameters for minimum performance targets that would allow for cross-industry comparison.

[464] Reddit response to November 2023 Illegal Harms Consultation, p.26; Mid Size Platform Group response to November 2023 Consultation, p.8. We note that Mid Size Platform Group made a similar point in response to the May 2024 Consultation on Protecting Children from Harms Online, p.9.
[465] Refuge response to November 2023 Illegal Harms Consultation, p.12.
[466] Institute for Strategic Dialogue response to November 2023 Consultation, p.9.

Different targets for illegal content versus content in breach of terms of service

A1.8.47 [✂].[467]

Our response

A1.8.48 The Act does not give us the power to recommend that service providers create targets for the removal of content that is not potentially illegal content. As outlined in paragraph 2.55 we have changed Measure ICU C4 to allow providers to set performance targets for content in breach of their terms of service (if they are satisfied that their terms of service are broad enough to cover illegal content). In measure ICU C5, we propose that services should consider the potential severity of harm (as well as the likelihood that content is illegal) when setting a policy for prioritising content for review.

Penalising excessively fast decision-making

A1.8.49 [✂].[468] techUK highlighted that moderation approaches have different timescales (with, for example, artistic or political content requiring a more deliberative approach).[469] It also highlighted that different content types have different review timescales, such as podcasts taking longer to review than images.[470]

Our response

A1.8.50 In the measure on performance targets, we recommend that providers balance the need to take content down swiftly with the importance of making accurate moderation decisions. As explained in paragraph 2.230, in Volume 2: chapter 2 'Content moderation', we consider that this recommendation offers important protections for users' rights, particularly their rights to freedom of expression. If a provider were to make moderation decisions so rapidly that it was taking down content inaccurately, then we consider this would have an adverse effect on users' rights.

Measure on a policy on the prioritisation of content for review

Marginalised or vulnerable groups

A1.8.51 Stakeholder feedback highlighted the importance of including the perspectives of marginalised or vulnerable groups when setting prioritisation frameworks. University College London (UCL) Gender and Tech Research Group suggested we consider these groups when determining what factors service providers should take into account in their prioritisation frameworks.[471]

Our response

A1.8.52 A diverse range of service providers fall within the scope of this measure, and the variation in risks across these services will affect different marginalised or vulnerable groups in

[467] [✂]
[468] [✂]; techUK response to November 2023 Consultation, p.8.
[469] techUK response to November 2023 Consultation, p.8
[470] techUK response to November 2023 Consultation, p.8
[471] UCL Gender and Tech Research Group response to November 2023 Illegal Harms Consultation, p.8.

different ways. It would therefore not be effective for us to prescribe a one-size-fits-all approach to engagement with marginalised groups in this measure.

Measure on resourcing content moderation

Risk assessments and resourcing

A1.8.53 Stakeholder feedback highlighted the importance of service providers adequately resourcing content moderation to consistently address harms according to their internal policies and performance targets. The NSPCC suggested that we should require providers to consider their risk assessment when resourcing moderation functions to ensure they have the necessary expertise to handle the most significant risk, specifically child sexual abuse and exploitation (CSEA).[472]

Our response

A1.8.54 We consider that the NSPCC's suggestion is already addressed by our recommendation that providers should resource their content moderation functions in accordance with their internal content policies. In our measure on internal content policies, we recommend that providers should have regard to their illegal content risk assessments when setting up these policies. In paragraph 2.140, in Volume 2: chapter 2: 'Content moderation' we explain that this means it would be reasonable to expect that a provider would include illegal harms with more than negligible risks on its service in its internal content policies. This should ensure that, when providers resource their content moderation functions in accordance with their internal content policies, this resourcing takes into account the service's risk assessment.

Make-up of content moderation teams

A1.8.55 Stakeholder feedback highlighted the importance of gender-inclusive moderation processes. Glitch suggested that without the involvement of women and girls in moderation processes, there will be gaps in understanding some content.[473]

Our response

A1.8.56 We do not consider it would be beneficial to specify how providers should resource their content moderation function. This includes the gender make-up of content moderation teams. However, we address Glitch's point about gaps in the understanding of individuals working in moderation in relation to specific kinds of illegal harms in paragraph 2.440 in Volume 2: chapter 2: 'Content moderation'.

Language

A1.8.57 Stakeholder feedback highlighted the importance of representing UK minority languages in content moderation and resourcing. The OSA Network expressed concern that there was

[472] NSPCC response to November 2023 Consultation, p.21. We note that the NSPCC made a similar point in its response to the May 2024 Consultation on Protecting Children from Harms Online, p.50.
[473] Glitch response to November 2023 Consultation, p.7. We note that VAWG Sector Experts made a similar point in its response to the May 2024 Consultation on Protecting Children from Harms Online, p.12.

no obligation for minority languages in the UK to be represented in content moderation functions, apart from Welsh.[474]

Our response

A1.8.58 As noted in Annex 4: 'Equality Impact Assessment and Welsh Language Impact Assessment', we are required to comply with Welsh language standards and must therefore consider the impact of our policy proposals in that regard.[475] However, under Measure ICU C6, depending on the UK userbase of the service, minority languages may also need to be considered when resourcing a content moderation function.

The impact of the measures on smaller, lower-risk services

A1.8.59 techUK suggested we conduct a further impact assessment on how resource allocation for content moderation affects smaller, lower-risk services and the implications of this for service quality.[476]

Our response

A1.8.60 Our measure on the resourcing of content moderation does not apply to providers of smaller, lower risk services.

Measure on the provision of training and materials to individuals working in content moderation (non-volunteers)

Contradictory language in the measure

A1.8.61 Stakeholder feedback highlighted unclear language in our November 2023 Consultation regarding content moderation. Logically suggested that our statements on the importance of content moderation training in our November 2023 Consultation were contradictory between paragraph 12.196 ("there may be occasions where training can be helpful in identifying and removing illegal content") and paragraph 12.209 ("the benefits of this measure are likely to be high…because training is important…to comply with…online safety duties").[477]

Our response

A1.8.62 In this instance, the stakeholder has taken the quote used in the November 2023 Consultation out of context. The full quote in paragraph 12.196 of the November 2023 Consultation is: "There may be occasions where harms-specific training and materials can be helpful in identifying and removing illegal content due to the unique, complex, novel or serious nature of a given harm, or because certain harm or harms may be particularly prevalent on a service and so require more in-depth understanding."

---

[474] Online Safety Advocacy Network response to November 2023 Illegal Harms Consultation, p.99.
[475] The Welsh language standards with which Ofcom is required to comply are available on our website. [accessed 5 November 2024].
[476] techUK response to November 2023 Consultation, p.3.
[477] Logically response to November 2023 Consultation, p.19.

Measure on the providing materials to volunteer moderators

Evidence for the measure

A1.8.63 An individual disagreed with the addition of the measure on providing materials to volunteer moderators to the Codes, and suggested there was not sufficient evidence provided for the measure.[478]

Our response

A1.8.64 We consider that we have provided sufficient evidence for this measure in the Codes, including evidence about the use of volunteers in content moderation, and the materials providers currently provide to volunteers.

Lack of vetting for volunteer moderators

A1.8.65 An individual expressed concern about the lack of vetting for volunteer moderators, showing particular concern that moderators will force their worldview on users and will be able to access other people's data and violate their privacy.[479]

Our response

A1.8.66 We do not consider it would be appropriate to provide guidance on the selection criteria for moderators, especially given that specific communities (rather than service providers) are most likely to be involved in the selection of volunteers.

[478] [✂]

[479] Name Withheld 3 response to May 2024 Consultation on Protecting Children from Harms Online, p.9.

A1.9 Codes of Practice: Search moderation

Stakeholder responses by theme

Scope of measures

A1.9.1 We received a number of stakeholder responses that argued our measures should apply to a different set of services. [✂] argued that regardless of their size, high-risk services should have obligations for internal content moderation policies tailored based on the frequency and severity of harms presented.[480] The Canadian Centre for Child Protection (C3P) argued that all search service providers should be in scope of measures ICS C4 and ICS C6. We note C3P also advocated for widening scope of the search moderation measures consulted on for our Children's Safety Codes in our May 2024 Consultation on Protecting Children form Harms Online ('May 2024 Consultation').[481]

Our response

A1.9.2 As outlined in our chapter on search moderation (Volume 2 chapter 3), we have assessed the impact of each measure and, based on this, have only recommended the search moderation measures to service providers when we consider it would be proportionate to do so. Our analysis suggests that it is proportionate to recommend all search service providers (including high-risk) apply measure ICS C1, and all providers of large general search services and multi-risk services apply measures ICS C2 to C6. We explain our reasoning in more detail in Volume 2, chapter 3 in the sections titled 'Who this measure applies to'.

Providers of small search services

A1.9.3 [✂].[482]

Our response

A1.9.4 The changes to measure ICS C1 set out in the 'Our decision' section remove references to technical actions (such as deindexing and downranking) that may not be applicable to providers of small search services. Clarifications set out in our explanation of Measure ICS C1 may also be helpful for providers of small services to understand expectations on them. We refer providers to our Illegal Content Judgement Guidance for guidance on assessing if content is illegal or not.

---

[480] [✂].
[481] Canadian Centre for Child Protection (C3P) response to November 2023 Illegal Harms Consultation, p.17; C3P response to May 2024 Consultation on Protecting Children from Harms Online, p.22.
[482] INVIVIA response to November 2023 Illegal Harms Consultation, p.12.

Moderation of content and impact on rights

A1.9.5 Oxford Disinformation and Extremism Lab (ODEL) noted that some content from civil society organisations, researchers, and human rights advocates risks being removed from or downranked in search results as a result of the operations of a search moderation function (for example, where extremism is discussed).[483]

A1.9.6 Big Brother Watch, Glitch, and Centro de Estudios en Libertad de Expresión y Acceso a la Información (CELE) expressed concerns about the impact of search moderation processes on access to information and freedom of expression. CELE added that, other than in the case of child sexual abuse material (CSAM), providers should not be making determinations on the illegality of content. Big Brother Watch made similar arguments in response to our May 2024 Consultation.[484]

Our response

A1.9.7 Our measures make recommendations as to how providers of search services can moderate illegal search content, which is in line with the duties placed on them under the Act and our duty to set Codes of Practice ('Codes'). We consider it proportionate for providers to take appropriate moderation actions in line with measure ICS C1 to minimise the risk of users encountering any illegal content they identify on the service. This is in line with the safety duties which apply to all priority illegal content and other illegal content that the provider is aware of.[485]

A1.9.8 While our measures relate only to moderation of illegal content, we recognise the risk of legal content being impacted by these processes, including as a result of incorrect identification of illegal content or moderation action being taken which impacts lawful content hosted alongside illegal content at a particular URL. We consider that our measures incorporate sufficient safeguards to mitigate against incidental effects on lawful content. For example:

- Measure ICS C1 does not recommend that identification of illegal search content should lead to it no longer appearing in results in all cases and recommends that providers consider a range of factors when deciding what moderation is most appropriate. This is to recognise the rights of users in relation to the legal content that is hosted at the same location as illegal content and which therefore could be affected by moderation.
- Measure ICS C3 recommends service providers balance the need to act swiftly in detecting and taking action in response to illegal search content with the importance of making accurate moderation decisions.

---

[483] Oxford Disinformation and Extremism Lab response to November 2023 Illegal Harms Consultation, p.10.
[484] Big Brother Watch response to November 2023 Illegal Harms Consultation, p.6; Big Brother Watch, May 2024 Consultation on Protecting Children from Online Harms, p.23; Glitch response to November 2023 Illegal Harms Consultation, p.7; Centro de Estudios en Libertad de Expresión y Acceso a la Información (CELE) response to November 2023 Illegal Harms Consultation, pp.7-8.
[485] Section 27(3) of the Act.

Scope of "search content"

A1.9.9 The Online Safety Act Network (OSA Network) noted that while content navigable via clicking on thumbnails was discussed in the 'Search' chapter of the Register of Risks ('Register'), the Codes do not extend to cover such content.[486]

Our response

A1.9.10 The Act defines "search content" as content that may be encountered in or via search results, and content will be treated as being encountered via search results if it can be accessed by interacting with search results (but does not cover content accessed following subsequent interactions).[487] References to "search content" reflect the scope of this definition. We have also amended paragraph 24.13 in the Register to reflect this. We agree with the wider risks highlighted by the OSA Network and have reiterated this in paragraphs 24.10-24.14 of the Register.

Harms-specific feedback

A1.9.11 [✂].[488] Glitch argued that greater consideration should be given to gender-based harm, including intersectionality, concerns around the speed at which intimate image abuse is removed, and the need for women in moderation teams.[489]

A1.9.12 An individual argued that any suicide or self-harm content not promoting recovery should be taken down, but with sympathetic support.[490] The Molly Rose Foundation expressed concerns that search engine algorithms not only recommend websites containing illegal suicide and self-harm content, but also prioritise them in search results.[491]

A1.9.13 Cifas argued that it would be easy and cost-effective to include known scam websites in deindexing by linking to free online scam website checker data.[492] [✂].[493]

A1.9.14 The Oxford Disinformation and Extremism Lab described how search functions continue to recommend content which is extremist but not branded as coming from a designated terrorist organisation. Therefore, the search moderation proposals should better account for the evolving nature of terrorist organisations and provide clear definitions of what is permitted and what is not.[494]

---

[486] Online Safety Act Network (OSA Network) response to November 2023 Illegal Harms Consultation, p.97.
[487] Section 57(2) and (5) of the Act.
[488] [✂].
[489] Glitch response to November 2023 Consultation, pp.6-7.
[490] Graham, Dr R. response to November 2023 Illegal Harms Consultation, p.2.
[491] Molly Rose Foundation response to November 2023 Illegal Harms Consultation, p.39.
[492] Cifas response to November 2023 Illegal Harms Consultation, pp.10, 13.
[493] [✂].
[494] Oxford Disinformation and Extremism Lab response to November 2023 Consultation, p.10.

A1.9.15 Glitch advocated for involvement of women in moderation teams given the prominence of gender-based harms.[495]

Our response

A1.9.16 Our first measure applies to all illegal content identified by providers of search services. This includes some of the harms areas raised by these stakeholders, such as terrorism, harassment, illegal suicide content, and fraud-related content. The search moderation measure recommends that providers take appropriate moderation action to minimise the risk of users encountering illegal content (with considerations given to severity of the content). Providers may consult our illegal content judgement guidance to help them make decisions in relation to whether search content is illegal.

A1.9.17 Our other measures will apply to services that have identified themselves as medium or high risk of at least two kinds of illegal harm (multi-risk). We would expect that the measures would therefore have an impact on reducing the risk of harm to these on their services.

A1.9.18 We do not consider there to be a strong enough case to change the search moderation measures to introduce additional or alternative harm-specific expectations or tools for detecting and moderating certain types of harms at this stage. We welcome feedback that we can take into account for future work in this area.

A1.9.19 In response to Glitch's point, we recognise the benefits of the involvement of women in moderation teams, but we continue to expect that providers are best placed to determine the resourcing and training needs of moderation teams based on the results of their risk assessments and addressing gaps in their moderator's understanding. Doing it on this basis will ensure that it is tailored to the reality of the individual service applying it and allow it to be more responsive to changes in the harms landscape. This could include addressing a gap in understanding on gender-based illegal harm.[496]

Moderator wellbeing

A1.9.20 Aside from the risks identified in our November 2023 Consultation, some stakeholders argued that performance targets put pressure on content moderators, leading to negative effects on moderator wellbeing such as burnout and mental health impacts. Zevo Health added that there could be a knock-on effect on user protection as welfare issues and staff turnover could impact how quickly and accurately moderators were able to remove illegal content.[497]

[495] Glitch response to November 2023 Consultation, pp.6-7.
[496] Glitch made this point in relation to search moderation, and we consider our comments are relevant to both search moderation and content moderation as well.
[497] Zevo Health response to November 2023 Illegal Harms Consultation, p.8.

Our response

A1.9.21 We acknowledge the possibility that moderator stress incurred from time-bound targets could lead to poor decision-making. However, we consider providers to be best placed to judge how moderator wellbeing may be impacted by performance targets and adjust them as needed. We are allowing providers flexibility regarding how to structure their targets, which will allow them to set targets that are conducive to wellbeing (if they consider that to be an issue for moderation staff).

Additional measures proposed

A1.9.22 We note that some additional measures were recommended to us, for example, from The Molly Rose foundation who proposed that search services should consider 'data voids' and take additional measures in response to the resulting risks.[498]

Our response

A1.9.23 As we are at an early stage in the regulatory regime, we are not considering whether additional measures in this area may be proportionate at this time.

[498] Molly Rose Foundation response to November 2023 Consultation, p.39.

A1.10 Codes of Practice: Automated search moderation

Stakeholder responses by theme

Support and recommendations

A1.10.1 We received several responses agreeing or in support with the measure, along with suggestions that we extend it to cover broader CSAM-related harms. Two stakeholders reflected on the prioritisation of existing measures and future developments of the Codes. Some of these responses are captured here.

Indecent imagery of children (IIOC hash lists)

A1.10.2 The [✂].[499]

Our response

A1.10.3 While indecent imagery of children (IIOC) hash lists are recommended as part of the measures for U2U services, we do not currently have sufficient evidence regarding their use in relation to search services (including the extent to which search services leverage hash lists in their automation).

Prioritisation of measures and additional measures

A1.10.4 Yoti and 5Rights Foundation argued for the prioritisation of different measures based on their impact on preventing child sexual abuse. They suggested that priority should be given to "preventative" measures – those that intervene further upstream in the perpetrator journey to prevent offences from occurring in the first place. They noted that while the measure would help prevent access to and viewing of URLs known to contain CSAM, this would happen after the point at which a child has suffered an act of abuse.[500] Specifically, 5Rights Foundation commented that "effective content moderation is an important *ex post facto* component of tackling harm online and is already used widely in the sector. However, it should not be prioritised over upstream measures to prevent illegal content and activity from being allowed to take place to begin with".[501]

A1.10.5 Yoti expressed disagreement with a lack of preventative features in our proposals, stating: "We do not think the approaches should be limited to three take down proposals, as detailed above, there should be equal focus on preventative measures… We strongly disagree that these takedown measures alone will be effective. We would invite Ofcom to

[499] [✂].
[500] 5Rights Foundation response to November 2023 Consultation, p.21; Yoti response to November 2023 Consultation, p.14.
[501] 5Rights Foundation response to November 2023 Consultation, pp.21, 23.

look at other prevention techniques currently available".[502] Yoti also noted a lack of costing discussion for such measures, stating: "We question as to why there is no inclusion of costing of preventative measures for CSAM in terms of costs?... It is worth considering the costs for law enforcement and civil society will continue to spiral, if preventative measures are not deployed across the ecosystem, with regulators working in conjunction with payment processors and ad networks as well as platforms."[503] These comments generally reinforce that removing results relating to listed CSAM URLs only help prevent access to CSAM and therefore offer only some alleviation of the harms caused by it.

Our response

A1.10.6 We acknowledge stakeholder suggestions that the measure should be accompanied by other measures with a greater focus on preventing the creation of CSAM, such as measures that deter perpetrators or protect children. As noted in our chapter 'Our Approach to developing Codes measures', our strategy is to publish measures quickly in order to not delay protections for users. As set out in Volume 2 chapter 5, through reducing the dissemination of CSAM, our measure to remove CSAM URLs from search results may contribute to prevention of child sexual abuse, given the correlation between viewing CSAM and perpetration of abuse. Similarly, our measure ICS E1, 2 and 3, as outlined in chapter 9, which recommends providers issue warnings to users seeking out CSAM to disrupt and deter their behaviour, and is applicable to all large general search services, also aims to prevent access to CSAM. Thus, at this time we have decided not to add any additional measures to what we consulted on in November 2023. However, we will continue to assess the proportionality of potential measures as we plan and work towards future iterations of our Codes.

[502] Yoti response to November 2023 Consultation, p.14.
[503] Yoti response to November 2023 Consultation, p.11.

A1.11 Codes of Practice: Reporting and complaints

Stakeholder responses by theme

Feedback on our approach

Alignment with the EU's Digital Services Act

A1.11.1 Some stakeholders highlighted the approach that the European Union's Digital Services Act ('DSA') has taken for equivalent provisions, with some suggesting closer alignment[504] and others suggesting greater flexibility to make it easier for service providers to harmonise their approaches across both jurisdictions.[505]

Our response

A1.11.2 Our review of the DSA's reporting and complaints duties show that there are some differences between our measure and the DSA's provisions, but the overall objectives of creating a safer space online are aligned. For example, the DSA requires services to communicate outcomes of reports to users, whereas at this time we are not in a position to determine that it would be proportionate for most services to introduce this, and recommend instead that services inform users of potential outcomes at the point of acknowledgement.

How our measures apply to downstream search services

A1.11.3 [✂].[506]

Our response

A1.11.4 Regarding [✂]'s concerns, see chapter 'Our approach to developing Codes measures'.

How our measures apply to private communications on certain encrypted services

A1.11.5 Safe Space One explained that, as a developer of end-to-end encrypted systems which are licensed to healthcare settings, it does not have access to the content and cannot be responsible for implementing the measures.[507]

[504] Mobile Games Intelligence Forum response to November 2023 Illegal Harms Consultation, p.1.
[505] Meta (WhatsApp) response to November 2023 Illegal Harms Consultation, p.2; Skyscanner response to November 2023Consultation, p.5; Wikimedia Foundation response to November 2023 Illegal Harms Consultation, p.27.
[506] [✂].
[507] Safe Space One response to November 2023 Illegal Harms Consultation, p.14.

Our response

A1.11.6 We recommend that providers seek independent legal advice regarding whether they are in scope of the Act and our measures.

A1.11.7 Providers of regulated services are required by the Act to operate complaints procedures. As set out in paragraph 2.57 of Volume 2: chapter 2: 'Content moderation', we consider that providers of end-to-end encrypted services in general have privacy-preserving ways to determine complaints about suspected illegal content, and that it is appropriate for them to do so. If the unique circumstances of Safe Space One enable it to secure that complaints about illegal content are determined in another way which protects its users equally well, under the Act it is open to take alternative measures.

Other, non-digital reporting and complaints procedures

A1.11.8 Some stakeholders argued that there should be routes to additional forms of support offline for complainants. University College London (UCL) Gender and Tech Research Group called for offline options to submit complaints for individuals whose devices might have been compromised.[508] Internet Matters called for Ofcom to consider requiring services to encourage children to speak to their parents or another trusted adult about reporting.[509]

Our response

A1.11.9 While we recognise the value of these suggestions, significant further work would be needed before we could make any such proposals, including careful consideration of whether they are matters which can appropriately be dealt with by the online safety regulator. Regarding Internet Matters' argument, we consider this to be covered by the point made in Volume 2: chapter 6: Reporting and complaints, under the section titled 'Submitting complaints as an 'affected' or 'interested' person or non-registered user.'

Reporting fake or misleading profiles

A1.11.10 In its response, Lovesaid brought attention to providers' lack of action against reported profiles.[510]

Our response

A1.11.11 As the Act's definition of 'content' is broad, we consider that 'user profiles' can be a form of content.[511] Therefore, a service provider should ensure that users are able to report the profiles of other users. In order to determine that a particular item of content is illegal content it may be necessary to consider other items of content. As set out in our Illegal Content Judgements Guidance, multiple items of content linked in this way may, depending on the circumstances, be illegal content. In particular, a service provider should ensure that users can report any illegal content that may be contained in the user profile or

[508] UCL Gender and Tech Research Group response to November 2023 Illegal Harms Consultation, p.10.
[509] Internet Matters response to November 2023 Consultation, p.17.
[510] Lovesaid response to November 2023 Illegal Harms Consultation.
[511] According to the Act: '"content" means anything communicated by means of an internet service, whether publicly or privately, including written material or messages, oral communications, photographs, videos, visual images, music and data of any description'. Section 236(1), Online Safety Act, 2023.

shared via the user's profile, including but not limited to financial scams. Where a provider does not provide a clear, easy-to-find, and accessible route to report, they have the right to complain about potential non-compliance with the Act to the service provider. We maintain that our measures on the appropriate action services should take will be sufficient to allow users to report fake or misleading profiles.

Measure on dedicated reporting channel for trusted flaggers for fraud (ICU D13/ICS D12)

Suggestions to expand our list of trusted flaggers in this measure

A1.11.12 In Volume 2: chapter 6: 'Reporting and complaints', we explained that several stakeholders made recommendations for how we might expand our list of trusted flaggers and the type of entities we might include in this expanded list. We have added two additional trusted flaggers to our specified list.

A1.11.13 We received a number of suggestions of other law enforcement and public sector organisations that should be named in the measure. We also received suggestions about naming other kinds of organisations as trusted flaggers in the measure, including specified anti-fraud organisations (SAFOs) such as Cifas, banking and financial services sectors bodies such as UK Finance, and other industry and civil society representatives.

Our response

A1.11.14 As explained in Volume 2: chapter 6: 'Reporting and Complaints', further expansion of the list of trusted flaggers named in the measure would depend on an assessment of the proportionality of doing so, including the costs and resource implications and the benefits to users.

Suggestions about expanding the measure to cover other harms and about how we choose trusted flaggers

A1.11.15 Several stakeholders, including civil society representatives, charities, public sector representatives and law enforcement, argued that dedicated reporting channels should not be fraud-specific but should be expanded to include other illegal harms, such as hate crime, child sexual abuse material (CSAM), violence against women and girls (VAWG), and firearms.[512] [✂].[513] One service noted that they already have dedicated reporting channels for trusted flaggers, which enable a range of reports, including fraud-related reports.[514]

[512] Name withheld – a civil society organisation response to November 2023 Illegal Harms Consultation, p.12; Refuge response to November 2023 Consultation, pp.16-17; Samaritans response to November 2023 Illegal harms Consultation, p.5; Scottish Government response to November 2023 Consultation, p.8; SWGfL response to November 2023 Consultation, p.15; UCL Gender and Tech Research Group response to November 2023 Illegal Harms Consultation, p.10.
[513] [✂].
[514] Meta response to November 2023 Consultation, p.29.

Another service noted that their trusted flagger programme is not limited to one type of harm but spans across and prioritises the most egregious illegal harms.[515]

A1.11.16 An individual respondent argued for greater transparency in the choice of trusted flaggers. This respondent argued that it is important to ensure that our choices did not perpetuate biases or a disproportionate focus on marginalised groups who may be disproportionately policed by law enforcement authorities, in particular highlighting how sex workers might be impacted.[516]

Our response

A1.11.17 We have not expanded the measure to include other harms at this stage because further clarity is required regarding the types of trusted flaggers that should be listed for those other harms. Further assessment is also required to understand the resourcing implications of listing trusted flaggers for other harms (such as VAWG and CSAM), specifically in relation to civil society and the voluntary sector. An extension of this measure may be considered in future iterations of the Codes. We recognise that transparency around the choice of trusted flaggers is important, particularly for establishing trust with marginalised communities. The only trusted flaggers we are currently recommending that service providers should establish relationships with are public bodies with expertise about fraud, and we have explained our reasoning in Volume 2: chapter 6: 'Reporting and complaints'. If we consider expanding our measure about dedicated reporting channels in the future to other kinds of trusted flaggers and/or harms, we would continue to consider carefully how the trusted flaggers should be selected.

Alternative information sharing arrangements

A1.11.1 Some stakeholders suggested automated data sharing (including the use of an application programming interface (API) and bulk data sharing) to improve the effectiveness of the measure and take further steps to tackle criminals involved in fraud.[517]

A1.11.2 UK Finance and Which? recommended the use of APIs to make intelligence sharing more efficient.[518] UK Finance recommended service providers should undertake information sharing with industries impacted by fraud (for example financial services sectors).[519] A stakeholder [✂] recommended bulk data dissemination to ensure the measure's effectiveness.[520] A civil society organisation proposed establishing a data sharing framework with a single common standard for such relationships.[521]

---

[515] Snap response to November 2023 Consultation, p.15.
[516] Are C. response to November 2023 Consultation, p.12.
[517] An API is a set of rules or protocols that enables software applications to communicate with each other.
[518] UK Finance response to November 2023 Consultation, p.7; Which? response to November 2023 Consultation.
[519] UK Finance response to November 2023 Consultation, p.7.
[520] [✂].
[521] Name withheld – a civil society organisation response to November 2023 Consultation, p.12.

A1.11.3 Cifas recommended reciprocal information sharing arrangements, noting that large services should also be encouraged to engage in multi-sector data and intelligence sharing mechanisms.[522]

Our response

A1.11.4 Currently, the measure does not specify exactly what types of information should be shared and how this should be done. This is for the service provider to determine jointly with each trusted flagger that requests access to a dedicated reporting channel. The type of information and how it should be reported should be discussed with the trusted flagger at the point of establishing the reporting relationship and at a minimum of every two years thereafter.

A1.11.5 We consider there to be value in encouraging reciprocal information sharing across multiple industries. However, reciprocal information sharing requires a clearer understanding of the cost implications for trusted flaggers. We encourage service providers to voluntarily engage in reciprocal information sharing arrangements.

[522] Cifas response to November 2023 Consultation.

A1.12 Codes of Practice: Recommender systems (U2U)

Stakeholder responses by theme

Offline testing

A1.12.1 The Integrity Institute suggested considering measures that recommend service providers carry out various types of offline testing.[523] Because offline testing is substantively different from on-platform testing, it would need to be considered as a new and separate measure in a future iteration of our Codes.

Our response

A1.12.2 We do not currently have sufficient evidence to suggest that offline testing methods are effective in evaluating the real-world illegal content risks related to design adjustments in recommender systems. Recommender systems are a type of artificial intelligence ('AI') that exhibit emergent behaviour influenced heavily by external stimuli (in this context, user behaviour, changes in content trends, and an evolving threat environment). Existing evidence suggests the most effective way of managing the risks of such a system is through continuous monitoring while live and operational.[524] On-platform testing allows service providers to observe the behaviour of systems in real time through user feedback and detect emerging patterns of risk.[525]

A1.12.3 Based on the research we commissioned, on-platform testing is one of the most advanced and effective methods for assessing the real-world risks of recommender systems.[526] However, we continue to develop our technical understanding of offline techniques to evaluate their effectiveness in helping service providers uncover the risks posed by their recommender systems. We will consider whether there is a case for incorporating offline testing to complement on-platform testing in future Codes in light of developments in this field.

Machine learning model transparency and third-party access

A1.12.4 The Integrity Institute suggested service providers should be required to publish the most important features of the machine learning ('ML') models used in their recommender

[523] Integrity Institute response to November 2023 Illegal Harms Consultation, p.17.
[524] Ofcom, 2023. Evaluating recommender systems in relation to illegal and harmful content. [accessed 8 October 2024]
[525] Ofcom, 2023. Evaluating recommender systems in relation to illegal and harmful content. [accessed 8 October 2024]
[526] Ofcom, 2023. Evaluating recommender systems in relation to illegal and harmful content. [accessed 8 October 2024]

systems to help users understand how the recommender system learns about each user and why it behaves the way it does.[527] [528] The Integrity Institute has suggested U2U service providers be required to make the safety metrics produced from on-platform testing in line with this measure publicly available for the use of third-party auditors.[529] The rationale is that this would ensure recommender system models undergo some level of external scrutiny and assurance.

Our response

A1.12.5 We recognise that publishing the most important ML features of a recommender system and related safety metrics (for example, via an API) may confer benefits to AI auditors and specialists by helping them evaluate and scrutinise why a recommender system might amplify illegal content. When published in an accessible format, such information may also help users better understand how their actions on a service might affect the content they encounter.

A1.12.6 However, we consider that there are risks and proportionality concerns associated with recommending the publication of such information as part of this measure. In particular, the information is likely to contain commercially sensitive and proprietary information and may be liable to be used by bad actors. As such, we do not consider it proportionate for us to recommend that providers publish any relevant design features or the safety metrics obtained by means of this measure.

Characterisation of recommender systems

A1.12.7 Airbnb and Booking.com agreed that recommender systems can amplify harm on certain services.[530] However, it also argued that the explanation of this risk factor should explicitly recognise that any risk is highly dependent on the purpose of the content shared on a service and the purposes for which a recommender system has been deployed. We have recognised that the risk of recommender systems varies according to their use case.

A1.12.8 Google said that we have over indexed the risk recommender systems and there has been insufficient acknowledgement of the benefits of recommender systems and of their importance in delivering a balanced and engaging user experience.[531]

---

[527] In machine learning, 'features' refers to measurable variables (such as user behaviours) that a recommender system observes to help it learn and make predictions about content the user is likely to engage with – this may include watch time, likes, comments, reshares, saves, scrolling habits, and more. 'Feature importance' involves understanding which features (such as user behaviours) have the most significant effect on a recommender system's outputs (so, what online behaviours does the model observe to make inferences about users' viewing habits, and which of these behaviours are given the most importance?)

[528] Integrity Institute response to November 2023 Consultation, p.18.

[529] Integrity Institute response to November 2023 Consultation pp.18-20.

[530] Airbnb response to November 2023 Illegal Harms Consultation, p.18; Booking.com response to November 2023 Illegal Harms Consultation, p.19.

[531] Google response to November 2023 Illegal Harms Consultation, pp.6-8, 57-59.

Our response

A1.12.9 We recognise the usefulness of recommender systems and their role in helping users encounter enjoyable content and potentially reducing the risk of harm. However, they can also increase the risk of harms, as described in the evidence presented in the Register. Given this risk, we think it is important to have a measure such as the one we have recommended.

Role of recommender systems in exposing children to harmful content online

A1.12.10 In response to measure ICU F1, two respondents said Ofcom's Illegal Content Codes of Practice do not capture the role of recommender systems in promoting harmful content to children. The Children's Commissioner noted that the measure does not cover recommender content (such as 'Discover' and 'For You' pages) based on users' interests and interactions. The Cyber Helpline also highlighted that we had not included push prompts for child users which encourage children to view content of a similar nature.[532]

Our response

A1.12.11 We have consulted on the risks posed by content recommender systems in relation to Primary Priority Content and Priority Content in the May 2024 Consultation on Protecting Children from Harms Online and have proposed measures in the Children's Safety Codes to prevent and protect children from being recommended content that it is harmful to them. We will publish our response to the May 2024 consultation in the first half of 2025.

Chat functionalities that allow multimedia sharing

A1.12.12 Evri highlighted that this measure does not apply to them as videos are sent directly from sender to recipient in a chat functionality. [533]

Our response

A1.12.13 We recognise that user generated content shared amongst users in a private chat environment has no direct relationship with content recommender systems, and therefore this measure does not apply to private chat functionalities.

[532] Childrens Commissioner's response to November 2023 Illegal Harms Consultation, p.23; Cyber Helpline response to November 2023 Illegal Harms Consultation, p.17.
[533] Evri response to November 2023 Illegal Harms Consultation, p.9.

A1.13 Codes of Practice: U2U settings, functionalities, and user support

Stakeholder responses by theme

Measure ICU F1 on safety defaults for child users

Age assurance accuracy & application

A1.13.1 The Cyber Helpline raised concerns about the accuracy and bias of age assurance technologies, which they felt are "still relatively unknown".[534] It suggested that age assurance methods alone cannot be used as effectively to determine whether a user is a child.[535] Yoti suggested that Ofcom should undertake research to assess the ease of circumvention of the full range of age assurance approaches.[536]

Our response

A1.13.2 We recognise that services use a variety of methods to determine a user's age or age range, which we discuss more fully in the section titled 'Determining a user's age or age range ' in volume 2, chapter 8: 'U2U settings, functionalities, and user support'. We continue to gather evidence to improve our understanding of the effectiveness such technologies and children's ability to circumvent age assurance methods.

A1.13.3 Our measure does not currently require the use of age assurance technologies for the purposes of service providers' implementation. Where services already have an existing means to determine a user's age or age range, they should use these for the time being to determine a user's age or age range for the protections under these measures. We discuss this more fully in the section titled 'How this measure works' in volume 2: chapter 8: U2U settings, functionalities, and user support. In this same section, we explain that once we publish our final decisions on our proposals on age assurance (as set out in our May 2024 Consultation), relevant service providers will, from that point onwards, need to use HEAA for the purposes of implementing the measures described in this chapter.[537] Under our

[534] The Cyber Helpline response to November 2023 Consultation, p.16.
[535] The Cyber Helpline response to November 2023 Consultation, p.16.
[536] Yoti response to November 2023 Consultation, p.15.
[537] As discussed in Volume 2: chapter 8: U2U settings, functionalities and user support, from the point at which the children's safety duties come into effect (which we expect around July 2025), relevant providers of U2U services will need to implement HEAA (where applicable) to comply with the children safety duties. Where relevant providers are using HEAA to comply with children safety duties, we expect that they will use HEAA to apply the measures in this chapter. Prior to this however, there will be a short period where service providers will have discretion as to what method they use to determine a user's age or age range for the purposes of implementing measures ICU F1 and F2.

proposal for HEAA set out in our May 2024 Consultation, services will be expected to consider Technical Accuracy and Fairness when implementing an age assurance process that is highly effective as part of our criteria based approach. This should address concerns about accuracy and bias in HEAA.[538]

Age verification

A1.13.4 One ID suggested that a broader range of services should be in scope of age verification requirements, which they also felt would align the Act with the ICO's Children's Code.[539] Refuge also raised a similar theme in relation to our measure on providing support to child users, suggesting that the measure should work alongside "robust" and "improved" age verification processes[540].

Our response

A1.13.5 As discussed in A1.4.3-4, the measures recommended in this chapter do not currently require the use of age verification or other age assurance technologies for the purposes of service providers' implementation. We also respond to this in the section below titled 'Overlap with the ICO's Children's Code'. However, from the point at which the Children's Safety Codes come into effect, relevant service providers of U2U services will need to implement HEAA (where applicable) to comply with the child safety duties and we expect these providers to use HEAA for the purposes of implementing the grooming measures in this chapter.

Enabling the identification of children

A1.13.6 Pinterest raised concerns that the recommendation to not display lists of users with whom a child on a service is connected could be used to identify children and therefore place them at risk.[541] It explained that on services such as Pinterest, where a user's age is not included in their profile, the absence of a 'Follower and Following list' (a connection list) would identify them as being under 18.

Our response

A1.13.7 We acknowledge that in limited circumstances the absence of a connection list could potentially result in children's accounts being easier to identify. However, we believe that the benefits from not displaying a child user account's connection list outweigh the potential risks arising from not displaying it. Furthermore, if potential perpetrators were to be able to identify child users in this manner and try to establish a connection, the supportive information measure helps mitigate against this risk. The supportive information measure recommends that child user accounts are provided with information on the types of interactions that would be enabled if they were to establish a connection

---

[538] See Ofcom, December 2023. 'Guidance for service providers publishing pornographic content. Consultation on draft guidance on age assurance and other Part 5 duties'.
[539] ICO, 2022. Age Appropriate Design Code: a code of practice for online services [accessed 31 October 2024]. We refer to this as the 'Children's Code'; One ID Ltd response to November 2023 Consultation, pp.2-3.
[540] Refuge response to November 2023 Consultation, p.19.
[541] Pinterest response to November 2023 Consultation, p.11.

and provides information on how they can take action against a user seeking to establish such a connection. Therefore, we consider this potential risk raised by the stakeholder to be minimal.

Cross-service offending

A1.13.8 The National Society for the Prevention of Cruelty to Children (NSPCC) and [✂] felt the measure did not account for the issue of cross-service offending (where perpetrators move their victims onto other services during the grooming process). The NSPCC and [✂] highlighted that this behaviour is an important part of grooming offences. [✂].[542] The NSPCC further suggested that perpetrators often use multiple accounts to evade children's attempts to block them or to stop engaging with them.[543]

Our response

A1.13.9 We acknowledge that cross-service offending is a significant issue and often forms part of the grooming process. We consider that the comprehensive nature of the illegal content risk assessment would likely bring the services that perpetrators are more likely to use for cross-service offending into scope of our recommended measures. Similar to the issue of the displacement of users (as we discuss in the section titled 'Who this measure applies to' in volume 2: chapter 8: U2U settings, functionalities, and user support), we consider that, if there is evidence of grooming occurring on the service to a significant extent, it is likely that these services would then be assessed as high risk. We would therefore expect the service to have to implement these measures. Providers need to be alert to any changes to their service and to update their risk assessment in line with our Risk Assessment Guidance.

Proactive detection of grooming

A1.13.10 While respondents generally supported and recognised the benefits of the measure to help tackle grooming, some felt that it was not enough in itself and that proactive solutions to detect grooming were also needed to help protect children from grooming risks.[544]

A1.13.11 The Scottish Government welcomed the measures, but felt that platform technologies could be better utilised to identify adults who make unsolicited contact with children (and who could be subsequently blocked).[545] Barnardo's and the NSPCC argued that the measure could be strengthened by targeting the behaviours of perpetrators (rather than those of children) by utilising proactive technology solutions to help identify grooming

[542] [✂]; NSPCC response to November 2023 Consultation, p. 18].
[543] NSPCC response to November 2023 Consultation, p. 18.
[544] Barnardo's response to November 2023 Consultation, p.2, p.11, p.21; NSPCC response to November 2023 Consultation, p. 35; Scottish Government's response to November 2023 Consultation, p.9; Snap response to November 2023 Consultation, pp. 19-20.
[545] Scottish Government response to November 2023 Consultation, p.9.

risks.[546] Barnardo's also noted that proactive solutions could take the form of machine learning and keyword detection to help identify grooming.[547]

Our response

A1.13.12 We acknowledge that the safety defaults measure is not a perfect solution. At the time of publishing the November 2023 Consultation, there was not yet enough evidence available for Ofcom to confidently recommend the use of proactive technologies to detect grooming. However, we are aware that the technology is developing and improving rapidly. We are monitoring developments closely and considering whether there is sufficient evidence to warrant including it in the Codes. If this analysis leads us to conclude that the technology is sufficiently mature, we will in due course consult on adding it to our Codes.

Additional settings and proposals for future measures

A1.13.13 Several stakeholders suggested additional safety settings or proposed future measures that would help prevent grooming.

A1.13.14 The suggestions included (but are not limited to):

- A suggestion for making children less discoverable online (such as removing personal information in a user's biography, or 'turning off' visible tags in shared content). Removing this personal information from a child's account ensures that such information is not visible to unconnected users.[548]
- A proposal to give children the ability to see who a direct message is from before deciding whether to accept or reject it.[549]
- A suggestion to ensure that children's accounts are set to 'private' by default, which would help to ensure that children are less easily identified online by perpetrators.[550]
- A suggestion that any livestreaming or in-game gifting functionalities should be 'turned off' by default, noting that this could help to ensure children are aware of livestreaming risks. As in-game gifting functionalities are often used by perpetrators to build a relationship with a child, such a setting could help ensure that children's susceptibility to commercial pressures through in-game gifting from perpetrators is reduced.[551]
- A proposal for network expansion functionalities allowing children to expand their network at a rapid rate to be 'turned off' by default, viewing such functionalities as encouraging children to add users they that do not know.[552]

---

[546] Barnardo's response to November 2023 Consultation, p.2, p.11, p.21; NSPCC response to November 2023 Consultation, p.25, p.35.
[547] Barnardo response to November 2023 Consultation, p.21.
[548] NSPCC response to November 2023 Consultation, p.38.
[549] NSPCC response to November 2023 Consultation, p.36.
[550] 5Rights response to November 2023 Consultation, p.28.
[551] 5Rights response to November 2023 Consultation, p. 25-28; CP3 response to November 2023 Consultation, p.24.
[552] 5Rights response to November 2023 Consultation, p.29.

- A suggestion for network expansion prompts that encourage users to engage with content of a similar nature, to be disabled as a default setting.[553]
- Proposals for additional safety measures, which are triggered if a child user changes a default setting, so additional friction is created to the grooming process. [✂].[554]

Our response

A1.13.15 As noted in 'Our approach to developing Codes measures', our strategy is to recommend measures quickly in order to not delay protections for users. Any additional measures would change our impact assessment and would require consultation. We do not have evidence at this stage that would justify such an approach but we will continue to keep this issue under review and as part of any future iterations of our Codes.

A1.13.16 Regarding proposals for additional measures if a child user seeks to change a default setting, we consider these issues are captured under measure ICU F2 (support for child users). If a child seeks to change a recommended setting or accept a direct message from another user for the first time, they will be provided with a supportive message informing them of the risks of doing so. This will help to reduce grooming risks and child sexual abuse and exploitation.

A1.13.17 We note the NSPCC's suggestion in response to the May 2024 Protecting Children from Harms Online Consultation ('May 2024 Consultation') that measure US1 (provide children with an option to accept or decline an invite to a group chat) should be extended to the Illegal Content Codes.[555] Any such extension would require additional evidence in relation to how this measure would reduce the risk and harms of grooming. We will continue to keep this issue under review.

Greater flexibility for services and discretion in implementing the measure

A1.13.18 WhatsApp suggested that service providers should be given greater flexibility regarding the implementation of the measure. It said that providers should be allowed to implement the safety defaults in a way that takes account of their existing safety measures, arguing that this would be particularly important for providers who have already "invested substantially" in protections for child users.[556] It further suggested that we should suggest or allow a range of options that would constitute "safe harbour".[557] WhatsApp also noted that a service provider should be treated as complying with the measure if it applies the "threshold" functionalities and default settings to all users, rather than needing to limit them to only child users or to otherwise provide a different experience to child users.[558]

---

[553] Cyber Helpline response to November 2023 Consultation, p.17.
[554] [✂]
[555] NSPCC Response to May 2024 Consultation, pp.65-66.
[556] WhatsApp response to November 2023 Consultation, p.31.
[557] WhatsApp response to November 2023 Consultation, p.16.
[558] WhatsApp response to November 2023 Consultation, annex, p.13.

Our response

A1.13.19 The default settings measure provides services with sufficient flexibility regarding its implementation. We have proposed measures which do not constrain providers unduly. However, for Codes to function as the Act requires, they need to be clear and detailed enough for it to be possible to say whether or not a provider has implemented the recommendations they contain. The Act provides flexibility for those who do not wish to adopt Codes measures to adopt alternative measures, so long as they record how this complies with the safety duty and their duties in relation to freedom of expression and privacy.

A1.13.20 We further encourage service providers to implement their safety settings alongside this measure where they feel that it strengthens the protection of children against grooming. We also address the issue of "safe harbour" as part of our 'Our approach to developing Codes measures'.

Overlap with the ICO's Children's Code

A1.13.21 5Rights noted that the measure appears to overlap with the Information Commissioner's Office (ICO) Children's Code.[559] It argued that Ofcom's Codes recommendations relating to the safety defaults should not "undermine" the default settings that service providers already apply to child users under the AADC by limiting the application of the measure to only those services with a high risk of grooming.[560] One ID raised a similar point and suggested that a broader range of services should be in scope of "age verification" requirements, which would align the Act with the AADC.[561]

A1.13.22 5Rights also suggested that while we develop our guidance around age assurance, we should align with the AADC in its application of the "likely to be accessed threshold", which determines the internet services within the scope of the AADC.[562]

Our response

A1.13.23 We have developed the measure with the knowledge that there are similarities between some of our measures and the standards of the AADC (for example, that geolocation sharing should be switched off). However, we note that the objective of this measure differs from that of the AADC. Our safety defaults measure focuses on mitigating grooming risks for children on services. In contrast, the AADC is focused on ensuring that service providers are protecting children's data online in accordance with data protection law.[563]

---

[559] 5Rights response to November 2023 Consultation, pp. 26; ICO, 2022. Age Appropriate Design Code: a code of practice for online services. [accessed 31 October 2024].
[560] 5Rights response to November 2023 Consultation, pp. 26. Regarding 5Rights' suggestion of the safety defaults measure appearing to overlap with the Children's Code, aspects of measure ICU F1 that were identified as being an existing regulatory requirement under the AADC are (1) that location sharing is switched off and (2) that children's data should not be disclosed unless there is a compelling reason to do so.
[561] One ID Ltd response to November 2023 Consultation. pp. 2-3.
[562] 5Rights response to November 2023 Consultation, pp. 20.
[563] ICO, 2022. Age Appropriate Design Code: a code of practice for online services. [accessed 15 October 2024].

Therefore, we expect the AADC to have a different approach in deciding which services fall within its scope.

A1.13.24 We also acknowledge 5Rights' suggestion that our proposals on scoping our measures' application should match the AADC's "likely to be accessed threshold" for determining which services are within scope of the AADC. However, as noted in paragraph A1.4.24, the different aims between our measures and the AADC result in different scoping methods. We also note that we have developed a risk assessment which considers the accessibility of a service by children – as well as other relevant factors – to determine the grooming risk level of a service and ultimately determine if that service is within scope of our measures. We discuss our proportionality considerations in the section titled 'Who the measure applies to' in volume 2: chapter 8: 'U2U settings, functionalities, and user support'.

Applicability of the measure to different messaging contexts

A1.13.25 The Integrity Institute expressed concerns that the safety defaults measure was not applicable to virtual reality (VR) technologies and other contexts where the messaging functionality is varied. It argued that the current wording regarding "receiving a direct message" was too prescriptive and failed to account for first-time interactions in future or niche contexts (such as VR) where messaging may not be the only method of contacting other users.[564] It suggested that the measure should be based on the principle of "this is the first time the child user is being contacted", without specifying the type or format of contact.

Our response

A1.13.26 We can confirm that the measure applies to all U2U service providers within scope, including VR services. We also note the Integrity Institute's concern about the wording "receiving a direct message", but we disagree that it is overly prescriptive. We considered many different forms of user communications, including communications across a range of services and technologies, as part of the measure's development. We identified that the direct messaging functionality is used widely across services and is a particular risk factor linked with grooming online. Additionally, expanding the safety defaults measure to other methods of contact would require further research to understand and analyse both the impact of the expansion on services and the efficacy of the measure. We recognise the need to consider evolving methods of contact online and will continue leading research to understand the different contexts (including VR) in which grooming may occur.

A1.13.27 We acknowledge the suggestion that the measure should be amended to include the principle of the first time that a child is contacted. However, we consider that, at this time, it would not be proportionate nor feasible to expand the measure without further research on the other methods of contact that would be in scope of the proposed expansion and without assessing the impact of such an expansion.

[564] Integrity Institute response to November 2023 Consultation, p.16.

Training and education

A1.13.28 Respondents from Sprite+ (University of Glasgow (UOG)) suggested that children needed training and education on how to use existing safety measures effectively (particularly for VR services) and argued for greater parental oversight of children's use of services.[565] Sprite+ (UOG) highlighted that VR was an environment with unique concerns and problems, which training on safety measures could mitigate. It suggested that we should consider how to improve "parental/guardian" insight and oversight of children's use of services.[566] INVIVIA and Snap also raised a similar theme of the role of parental tools in managing children's safety online.[567]

Our response

A1.13.29 We recognise the challenges and unique threats that VR technologies present to children's safety. We are working to better understand these threats in a VR context to help support our assessment and analysis.

A1.13.30 We acknowledge that service providers have a role to play in educating parents, guardians, and children about the risks of grooming and relevant safety measures. However, there is limited research on the effectiveness of parental controls across different service types to mitigate against the risks of child sexual exploitation and abuse (CSEA), especially grooming. We also are aware of the importance of recommending parental tools that increase safety and reduce the potential for misuse of services. However, we did not receive substantive evidence from respondents that would justify the inclusion of either greater parental or guardian oversight or specific training on safety measures as part of default settings measure. We will continue to keep this issue under review and will consider revisiting our analysis if (and when) we are presented with such evidence.

Measure ICU F2 on support for child users

Measure placing responsibility on children and not perpetrators

A1.13.31 Some stakeholders raised concerns that the measure places too great an emphasis on children keeping themselves safe, rather than focusing on targeting perpetrators and preventing them from engaging in illegal behaviours.[568]

---

[565] Sprite+ (University of Glasgow (UOG)) response to November 2023 Consultation, pp.14-15. We note that REPHRAIN made a similar point in May 2024 Consultation on Protecting Children from Harms Online, pp. 20-21.

[566] Sprite+ (UOG) response to November 2023 Consultation, pp. 14-15.

[567] INVIVIA response to November 2023 Consultation, p.20; Snap response to November 2023 Consultation, p.20. We note that Snap made a similar point in the May 2024 Consultation, pp. 12, 25-26. This feedback however, while raising a similar issue, was in relation to the measure we refer to as US6 (Provide age-appropriate user support materials for children), and Ofcom's consideration of parental controls as an area for potential future consideration.

[568] Lucy Faithfull Foundation response to November 2023 Consultation, p.10; Welsh Government response to November 2023 consultation, p.4. WeProtect Global Alliance response to November 2023 consultation, p.20

Our response

A1.13.32 This measure is designed to empower children to make informed choices regarding the types of content that they view and the users they interact with online. This measure therefore makes it harder for perpetrators to identify and interact with children. It will be a significant improvement on the status quo and likely lead to the reduction of CSEA.

Supportive messaging as part of reporting process

A1.13.33 While indicating broad support for the measure, Nexus argued for the need for accessible and robust complaints and reporting process for children, particularly in relation to illegal or harmful content.[569] To achieve this, they suggested providing supportive messaging after a complaint is made, explaining what happens to an account during this process (for example, suspension pending investigation). NWG Network also raised a similar point, noting that the measure should ensure that there is a consistent reporting pathway and a feedback loop for children when they report abuse.[570]

Our response

A1.13.34 We agree that service providers should ensure that their reporting processes are clear and accessible. While many providers currently have reporting tools or complaints processes in place, our evidence suggests that many people (particularly children) face barriers to using them.[571] Problems include reporting tools being difficult to find or not clearly identifiable and complaints systems being too burdensome or involving too many steps.[572] This measure (support for child users) helps address some of these barriers by informing child users about their options to report conduct when they take action against a user or before they accept a direct message from a user for the first time.

A1.13.35 Our other measures also directly address this, specifically those discussed in volume 2: chapter 6: 'Reporting and complaints'. These measures require all service providers to design and operate complaints systems and processes so that they are easy to find, easy to access and easy to use. Our measures also require some service providers to provide additional information on the outcome of a complaint and how long it will take to handle the complaint. Please see our measures outlined in volume 2, chapter 6 for further detail.[573]

---

[569] Nexus response to November 2023 Illegal Harms Consultation, p.15.
[570] NWG response to November 2023 Illegal Harms Consultation, p.9.
[571] Ofcom, May 2024. Protecting Children from Harms Online, Vol 5: What should services do to mitigate risks?, p.224-225. [accessed 22 November 2024].
[572] Ofcom, 2024. Understanding Pathways to Online Violent Content Among Children. [accessed 15 October 2024].
[573] See Volume 2: chapter 6: Reporting and Complaints.

Additional user points where supportive messaging is needed

A1.13.36 Several stakeholders suggested additional points for supportive messaging.[574] Suggestions included (but are not limited to):

- A supportive message displayed to a user at the point of the sign-up process for the account's creation.[575]
- A supportive message displayed to a user when there are reminders about any new or updated settings.[576]
- A supportive message displayed to a user during the "age verification process" noting what to do in cases of sharing or accessing illegal or harmful content and highlighting the potential dangers of doing so.[577]
- A supportive message displayed to users when illegal content is removed, explaining the reason and the potential risks.[578]
- A supportive message displayed to a user who seeks to change their default setting to share their precise location with another user, with a message displaying the potential risks of sharing such information.[579] If a child 'turns off' a setting, they could also receive ad hoc messages to review their settings (this could be displayed annually or "at the end of every season", once a user has a certain reach, or when they engage with a new feature).[580]
- A supportive message displayed to a user before engaging in livestreaming explaining the risks involved (for example, that the livestream could be recorded by another user).[581]
- A supportive message which 'flags' risks or restrictions to interactions based on factors beyond those taken by an individual user (such as warning messages regarding someone who has been blocked or reported by other users or is from a region where the child's network is not typically located).[582]

---

[574] 5Rights response to November 2023 Consultation, pp.27-29; C3P response to November 2023 Consultation, p24; NSPCC response to November 2023 Consultation, p.37-38; WeProtect response to November 2023 Consultation, pp.21-22.
[575] NSPCC response to November 2023 Consultation, p.38.
[576] NSPCC response to November 2023 Consultation, p.37.
[577] We infer "age verification process" as the process through which a user may confirm their age; WeProtect response to November 2023 Consultation, p.21.
[578] WeProtect response to November 2023 Consultation, p.21
[579] Snap response to November 2023 Consultation, p. 21.
[580] NSPCC response to November 2023 Consultation, p.38; 5Rights response to November 2023 Consultation, p.30.
[581] C3P response to November 2023 Consultation, p.24.
[582] We note that Snap made a similar point in the May 2024 Consultation on Protecting Children from Harms Online, p.26, regarding measure US4 (Provision of information to child users when they restrict interactions with accounts or content). This feedback, while raising a similar issue, has a different context compared to

A1.13.37 As highlighted in the relevant 'Effectiveness' section in volume 2: chapter 8: 'U2U settings, functionalities, and user support', it is important to reach children at the right time and with the right messaging while also avoiding prompt fatigue (which could result in information prompts becoming less effective). Additionally, repeated supportive information or additional supportive information can be materially costly for service providers to design and implement, as well as imposing indirect costs on service providers through reduced user engagement (see the relevant 'Costs and risks' section for this measure).

A1.13.38 Furthermore, as we note in chapter 8, children would also be provided with a supportive message if they seek to 'turn off' one of the safety defaults. The information provided should assist children in understanding the implication of making this change, including the protections afforded by the setting that they wish to disable.

A1.13.39 As discussed in 'Additional settings and proposals for future measures', any additional measures would change our impact assessment and would require consultation. We do not have evidence at this stage that would justify such an approach.

Informing users of the risks of sharing illegal content & indecent images

A1.13.40 The New Zealand Classification Office suggested that service providers would need to strike a balance between informing younger users of the risks of processing or sharing illegal content, against the risk of children then deciding to share this content with other users.[583]

A1.13.41 The NWG network highlighted that Ofcom could also improve the supportive messaging to address the harm of children sending and receiving indecent images. It noted resources from NWG, Childline, Marie Collins Foundation, the Lucy Faithfull Foundation, and the NSPCC, all of which could provide support to affected families and young people.[584]

Our response

A1.13.42 The support for child users measure (measure ICU F2) will mitigate the risks of children encountering illegal harms with a specific focus on grooming for the purposes of CSEA. The measure is designed to increase the availability of supportive information to children, to help them make informed choices about risk in their online experience. It is not designed to specifically inform child users of the risks of sharing illegal content or indecent images. It is more focussed on informing child users about the risks of using particular functionalities and how to protect themselves if necessary. In any event, we expect providers to consider in their design of the supportive information, the balance of the information provided to child user accounts so that it helps them make informed choices.

---

measure ICU F2 (support for child users), which requires a supportive message at the point where a child takes action against another user's account, compared to US4 which requires such messages when child users restrict interactions with other accounts **or content.** See paragraphs 8.207 to 8.209 for more details.

[583] New Zealand Classification Office response to November 2023 Consultation, p.9.

[584] NWG response to November 2023 Consultation, p.8.

A1.13.43 We note that in relation to supportive information on the risk of sharing content, our proposed Children's Safety Codes recommend user support measures to inform children of the risks of searching for content that is harmful to children (such as suicide, self-harm and eating disorder content), or posting and re-posting harmful content.[585] We will publish our decision on these proposed measures in 2025.

[585] US5 signposts children to support at key points in the user journey, including intervention point 2 (when children post or re-post content) and intervention point 3 (when children search for harmful content). For further details, please see: Ofcom, May 2024. 'Protecting Children from Harms Online'.

A1.14 Codes of Practice: Search settings, functionalities, and user support

Stakeholder responses by theme

Provision of CSAM content warnings

Setting up and updating the list of CSAM search terms

A1.14.1 Two civil society organisations commented on the process for setting up and updating the list of terms used to trigger a CSAM warning under the CSAM warning measure.

A1.14.2 C3P commented on the need for a "source of data" that provides the keywords and symbols in the list of terms used to trigger the CSAM content warning.[586] Both C3P and Protection Group International noted that the list of CSAM terms should be reviewed and updated on a regular basis as perpetrators develop new terminology to evade detection.[587]

Our response

A1.14.3 We consider that the measure addresses these concerns. Regarding the suggestion around the source of data, the measure recommends that an appropriate list of search terms should be developed and maintained by (or sourced from) a person (or organisation) with expertise in the terms commonly used to search for CSAM. Regarding the second suggestion around reviewing the list, the measure recommends that the list should be regularly updated to add and remove terms as necessary. We therefore conclude that the measure recognises and addresses the risk of the language used by perpetrators evolving.

A1.14.4 Protection Group International sought clarification as to whether the list should cover terms related to CSAM only or should include terms more broadly related to CSEA offences.[588] The measure that we have recommend in the Code only covers search terms relating to CSAM offences. Providers may wish to cover a wider range of CSEA offences as a matter of their own user safety practices.

A1.14.5 Schedule 6 of the Act sets out a number of priority offences relating to CSEA. For the purpose of presenting our analysis, the Register sets out two broad categories of CSEA

[586] C3P response to November 2023 Consultation, p.29.
[587] C3P response to November 2023 Consultation, p.29; Protection Group International response to November 2023 Consultation, p.13.
[588] Protection Group International response to November 2023 Consultation, p.13.

offences: CSAM and grooming.[589] We recognise that these two categories of CSEA are often closely related and may happen either in isolation, in parallel, or sequentially. However, their separation in the Register reflects important differences in the type of offence and the online settings in which they may occur.

A1.14.6 A full list of priority offences included in the CSAM category is set out in the Register chapter titled 'Child Sexual Exploitation and Abuse (CSEA)'. These mainly include offences involving images or videos depicting sexual acts with children, any other indecent or prohibited imagery of children (including 'pseudo-photographs' that have been made entirely on a computer), and paedophile manuals.[590] The grooming category includes offences relating to sexual communication and interaction with a child to facilitate online or offline child sexual abuse.

A1.14.7 Our assessment of the risk of harm on search services found they provide perpetrators with a frequently used and easy point of access for finding CSAM. We have not seen similar evidence of search services facilitating behaviour-based grooming offences.[591] Therefore, we consider it appropriate to limit the scope of the measure to search queries that indicate that a user is seeking to encounter CSAM. This is also likely to capture search queries that a grooming perpetrator might enter when using a search service.

Provision of suicide crisis prevention information

Extension of crisis prevention information to other harms

A1.14.8 Two stakeholders called for service providers to provide crisis prevention information for other harms under this measure, namely fraud and self-harm.[592]

Extension to fraud

A1.14.9 The FCA highlighted that service providers could display fraud prevention messaging when users search for financial services or products. It suggested this messaging might lead users to pause or seek help before potentially losing money to fraud.[593] There is evidence of banks already using positive frictions in this context – for example, a 2023 report by UK Finance indicates that the use of a warning helped contribute to a reduction of authorised payment push fraud.[594]

Our response

A1.14.10 We recognise that there may be potential benefits to service providers displaying fraud prevention messaging in response to searches for certain financial products and services

[589] See the Register chapter titled 'Child Sexual Abuse and Exploitation (CSEA)'.
[590] See the Register chapter titled 'CSEA'.
[591] See the Register chapter titled 'CSEA'.
[592] FCA response to November 2023 Consultation, pp.6-7; Samaritans response to November 2023 Consultation, p.5.
[593] FCA response to November 2023 Consultation, pp.6-7.
[594] FCA response to November 2023 Consultation, pp.6-7; UK Finance, 2023. 2023 Half Year Fraud Update. [accessed 18 November 2024].

such as debt advice or high-risk investments. However, further evidence is required to understand the effectiveness of displaying fraud prevention messaging, as recommending this would constitute a more material change in search service providers' existing safety procedures. If we decide to develop a measure on this theme in future, we would want to ensure it is proportionate and can effectively target the relevant risks. We would therefore need to gather further evidence on the risk of encountering fraudulent financial products and services via search services.

Extension to self-harm

A1.14.11 Samaritans argued that service providers should display crisis prevention information where users search for terms relating to self-harm ideation or to methods of self-harm.[595] The Molly Rose Foundation cited one of our own studies indicating that users were six times more likely to encounter harmful content about self-injury when entering deliberately obscure search terms.[596]

Our response

A1.14.12 The crisis prevention measure is designed to address the risks associated with users encountering search content that amounts to the priority offence of encouraging or assisting suicide.[597] While the Act also introduces the offence of encouraging or assisting serious self-harm, this is not a priority offence.

A1.14.13 We acknowledge the importance of offering timely assistance to users who are experiencing mental health concerns involving self-harm and who conduct searches that are likely to lead them to harmful content. That said, Parliament's decision to define certain offences as priorities in the Act suggests that illegal content classified in this way should be treated more seriously. At this early stage in the establishment of the regulatory regime, we do not consider it proportionate for us to recommend that providers build their systems and processes in a way that enables them to consider all potentially relevant non-priority offences as well as priority offences.

A1.14.14 While we do not consider self-harm to be directly within the scope of this measure, we have proposed another measure that does relate to self-harm in a different Code and under a more relevant duty in the Act. Content that encourages, promotes, or provides instructions for self-harm is a category of primary priority content covered by the child safety duties that apply to search service providers under section 29 of the Act.[598] In the May 2024 Consultation, we proposed a measure recommending that providers of large

---

[595] Samaritans response to November 2023 Consultation, p.5.

[596] These include abbreviations, insertions of symbols into real words, or the use of homonyms. Source: Molly Rose Foundation response to November 2023 consultation, p.39; Network Contagion Research Institute, 2024. One Click Away: A Study on the Prevalence of Non-Suicidal Self Injury, Suicide, and Eating Disorder Content Accessible by Search Engines. [accessed 18 November 2024].

[597] The offence of encouraging or assisting suicide under section 2 of the Suicide Act 1961 and section 13 of the Criminal Justice Act (Northern Ireland) 1966 are included as priority offences under paragraphs 1 and 2 of Schedule 7 to the Act.

[598] See the definition of "primary priority content" in section 61(4) of the Act.

general search services likely to be accessed by children should provide crisis prevention information in response to search requests regarding self-harm (alongside suicide and eating disorders).[599] We proposed that the measure in the Childrens' Safety Codes should be applied to all logged-in and logged-out search users. In practice, we expect that measure to cover the same services as this measure in the Codes, and we consider the risks associated with users encountering self-harm content should be addressed by that more relevant measure.

Nation-specific signposting

A1.14.15 In response to the crisis prevention measure we proposed in the May 2024 Consultation, the Scottish Government suggested some Scottish-based websites for providers to signpost to as part of their crisis prevention information.[600]

Our response

A1.14.16 While we recognise the benefits of signposting users to nation-specific resources, we do not consider it proportionate at this stage for us to recommend a measure that would, in practice, require providers to know the precise location of its users in order to signpost in such a targeted manner. There may also be limited nation-specific helplines that are available 24/7.

A1.14.17 We nonetheless consider that the measure, which recommends that providers signpost to a 24/7 helpline that is available to all UK users and is suitable for all ages, will still offer important benefits by directing users towards resources that are relevant and available to users across the UK. It remains open to a service provider to provide additional nation-specific resources as part of its crisis prevention offering where relevant and technically feasible to do so.

Suicidal ideation

A1.14.18 CarefulAI expressed concern that we have not recommended service providers should "look for and act upon suicidal ideation language".[601]

Our response

A1.14.19 We expect providers to display crisis prevention information in response to terms expressing suicidal ideation. As explained in the chapter at paragraph 9.169, the measure sets out that service providers should detect and provide crisis prevention information in response to search requests relating to suicide. In recognition of the fact that suicidal ideation is a spectrum, the measure covers both general queries regarding suicide and queries seeking specific, practical, or instructive information regarding suicide methods. This should ensure that the search journeys of vulnerable users are disrupted whether they are speculatively browsing for suicide content or more purposefully searching for detailed information on suicide methods.

---

[599] See Measure PCS E3 in Annex 8 in May 2024 consultation. [accessed 18 November 2024].
[600] Scottish Government response to May 2024 Consultation, p.19.
[601] CarefulAI response to November 2023 Illegal Harms Consultation, p.1.

Non-measure specific responses

Data voids

A1.14.20 The Molly Rose Foundation highlighted the risk of "data voids" and requested that we require services providers to take additional measures in response to those risks.[602]

A1.14.21 As defined in our 2024 research on the on the Prevalence of Non-Suicidal Self-Injury, Suicide, and Eating Disorder Content Accessible by Search Engines, data voids are "situations where the search demand for certain keywords is not met with reliable or safe information due to relative obscurity of search terms of phrases".[603]

A1.14.22 The Molly Rose Foundation noted that data voids increase the risk of search terms that contain cryptic language leading users to more harmful content "as algorithms aim to provide relevant results but lack necessary safe and accurate information to fill those gaps".[604]

Our response

A1.14.23 We do not consider this risk to be directly relevant to our measures relating to CSAM warnings and crisis prevention information beyond providing further evidence for the risks posed by evolving language. We would expect providers to keep terms that trigger crisis prevention under review to ensure effectiveness of this intervention. We will continue to build our understanding of the risks posed by data voids and consider how we might address these in future updates to the Codes.

Fines

A1.14.24 The NSPCC suggested we consider how we could use income generated from fines levied under the Act to fund the support organisations cited in the measures relating to CSAM warnings and crisis prevention, particularly where those organisations provide helplines that "require significant resources".[605]

Our response

A1.14.25 Under section 400 of the Communications Act 2003 (as amended by section 206 of the Act), we must pay any amount paid to us in respect of a penalty imposed or any additional fee charged under the Online Safety Act 2023 into the Consolidated Fund of the United Kingdom. Therefore, we do not have discretion as to how those funds are used.

Encountering harmful content within one click of the search results page

A1.14.26 The Online Safety Act Network highlighted areas in the Register where we noted the risks of encountering harmful content within one click of the search results page (for example,

---

[602] Molly Rose Foundation response to November 2023 consultation, p. 39.
[603] Network Contagion Research Institute, 2024. One Click Away: A Study on the Prevalence of Non-Suicidal Self Injury, Suicide, and Eating Disorder Content Accessible by Search Engines. [accessed 18 November 2024].
[604] Molly Rose Foundation response to November 2023 consultation, p. 39; Network Contagion Research Institute, 2024. One Click Away: A Study on the Prevalence of Non-Suicidal Self Injury, Suicide, and Eating Disorder Content Accessible by Search Engines. [accessed 18 November 2024].
[605] NSPCC response to November 2023 Consultation, p.45.

clicking through thumbnails to access harmful content). It highlighted that there is no mention in the Codes of the "one-click limit" or of how service providers can mitigate the risks described in the Register.[606]

Our response

A1.14.27 Our measures focus on interventions that anticipate and address the risk of encountering search results containing illegal content. In chapter 3 of Volume 2: 'Search moderation', we explain that the Act defines "search content" as "content that may be encountered in or via search results. This extends to content that can be accessed by users directly through interacting with search results (for example, by clicking on them) but does not extend to any content that can be encountered because of subsequent interactions.[607] Our measures cover search settings, functionalities, and user support and do not recommend any action to be taken regarding specific pieces of search content. We discuss the "one-click limit" where relevant in paragraphs 9.23, 9.104, and 9.193.

GenAI-powered search services and violence against women and girls ('VAWG')

A1.14.28 Refuge expressed concerns around the potential growth of new search engines driven by GenAI that focus on identifying individuals and that might have "dedicated harmful purposes." It provided some examples of search services and noted how these may make it easier for perpetrators of VAWG to find information about victims and survivors.[608]

Our response

A1.14.29 Regarding the specific examples of services provided by Refuge, it is for the provider of a search service to determine in the first instance whether that service is in scope of the Act. Where a search service is in scope of the Act, providers need to comply with the relevant illegal content safety duties. This includes conducting a risk assessment against their features and functionalities (such as the use of Generative Artificial Intelligence ('GenAI').

A1.14.30 Regarding Refuge's concern about the emergence of GenAI-powered search services with functionalities dedicated to causing harm in future, we have established projects exploring the implications of GenAI for online safety. This includes exploring how providers are integrating GenAI to power aspects of their services, how GenAI can enable online harm, and how providers can respond to mitigate these harms. To date, we have published two papers on red teaming (a type of AI evaluation) and interventions to address harmful deepfakes.[609] We are pursuing an iterative approach to our Codes as we learn more about these harms and mitigations.

---

[606] Online Safety Act Network response to November 2023 Illegal Harms Consultation, p.97.
[607] Section 57(2) and (4) of the Act.
[608] Refuge response to November 2023 Illegal Harms Consultation, p.15.
[609] Ofcom, 2024. Red teaming for GenAI Harms – Revealing the Risks and Rewards for Online Safety. [accessed 18 November 2024]; Ofcom, 2024. Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes. [accessed 18 November 2024]. We also published an open letter about how the Act will apply to Generative AI and chatbots in November 2024. Ofcom, 2024. Open letter to UK online service providers regarding Generative AI and chatbots. [accessed 18 November 2024]

Applicability of measures to GenAI features

A1.14.31 In the May 2024 Consultation, we proposed counterparts to the measures relating to predictive search and crisis prevention information outlined in this Statement, focusing on content harmful to children.[610] We asked stakeholders if they considered it technically feasible to apply those measures to GenAI functionalities likely to perform or be integrated into search functions.

A1.14.32 The NSPCC responded that it considers it both technically feasible and "highly desirable" for providers to apply those measures to their GenAI functionalities.[611] One individual [✂] disagreed with this stance.[612] While the Centre for Excellence for Children's Care and Protection (CELCIS) did not express a view on technical feasibility, it expressed support for any measures that might address the risk of AI-generated CSAM.[613]

A1.14.33 Microsoft noted that search services with GenAI-integrated features require distinct measures. It proposed that measures should include recommendations for the use of metaprompts and classifiers to mitigate risks of generating or encountering harmful content on GenAI-enabled search services. It also suggested that GenAI search functions should provide links and citations for information to reduce the risk of misinformation and disinformation.[614]

Our response

A1.14.34 We have not seen sufficient evidence at this stage to support or inform introducing measures relating to search settings, functionalities, or user support for GenAI-enabled search services. We will keep this under review.

---

[610] Measure SD1 proposed that providers should offer users a means to easily report predictive search suggestions relating to primary priority content (PPC) and priority content (similar to measure ICS F1). Measure SD2 proposed that providers should display crisis prevention information in response to known PPC-related search requests regarding suicide, self-harm, and eating disorders (similar to measure ICS F3).
[611] NSPCC response to May 2024 Consultation, p.71.
[612] [✂]
[613] CELCIS response to May 2024 consultation, p.19.
[614] Microsoft response to May 2024 Consultation on Protecting Children from Harms Online, p.13.

A1.15 Codes of Practice: Terms of service and publically available statements

Stakeholder responses by theme

Prompts

A1.15.1 As set out in our November 2023 Consultation, we considered evidence suggesting that 'prompts' (information provided to users in a brief and timely way) around terms of service ('terms') and publicly available statements ('statements'), or changes to terms and statements, could improve user understanding. We provided some examples of evidence we received in response to our 2022 Illegal Harms Call for Evidence. We also raised concern over 'prompt fatigue' and the need to justify added friction for users. We asked stakeholders to include relevant information on prompts in their response to our November 2023 Consultation question.

A1.15.2 Match Group provided examples of how prompts have been used on its service to build understanding and shape user behaviour by 'warning' users to reconsider potentially offensive, harassing, or problematic language.[615] Oxford Disinformation and Extremism Lab cautioned that prompts are "only one part of the overall solution as they may be ignored or disregarded" (particularly when users experience greater exposure to them).[616] It also emphasised the importance of incorporating behavioural psychology insights, empathy, and humour into the development of prompts, and the value of ensuring that prompts and other 'nudge' methods are continually updated and changed to ensure they are considered credible by users.

Our response

A1.15.3 We value the feedback on prompts and will continue to consider the use of prompts in future possible iterations of the Codes. However, we have insufficient evidence at this stage to systematically evaluate the effectiveness of these types of prompts across all services in-scope of the Act.

A1.15.4 Our Behavioural Insights Hub has conducted behavioural trials into the use of prompts to improve engagement and comprehension of terms of service.[617] The research suggested that while prompts can improve engagement with terms of service, they are unlikely to have a dramatic effect on levels of user engagement on their own. Therefore, it is

[615] Match Group response to November 2023 Consultation, p.14.
[616] Oxford Disinformation and Extremism Lab response to November 2023 Consultation, p.15.
[617] Ofcom, 2024. Promoting user engagement with Terms and Conditions. [accessed 29 November 2024]

important for service providers to consider other strategies for encouraging user engagement with terms of service rather than relying only on general prompts. This may involve providing reminders of key information about terms and conditions at other salient points in the user journey (such as uploading or reposting content).

A1.15.5 In this iteration of the Codes, we are not recommending a specific measure regarding the use of prompts to encourage or increase user engagement with terms and statements. However, the use of prompts as a tool to support user empowerment more generally is an area we propose to continue to research and explore, and it is possible that we will come back to this issue as new evidence or research emerges.

A1.15.6 We also invite stakeholder expressions of interest to collaborate with our Behavioural Insights Hub on testing the use of prompts in different contexts and at different stages in the user journey to help develop the evidence base in this area.

The role of Ofcom

A1.15.7 The Federation of Small Businesses suggested Ofcom consider providing a terms and statements generator where businesses can input information to be customised to their needs, helping to alleviate any uncertainty around what information to include and making it easier to comply with the measure.[618]

A1.15.8 5Rights Foundation suggested we map service providers' terms and statements over time to track how they are making changes and to ensure they are meeting the spirit of the Act.[619]

A1.15.9 Snap queried the absence of greater alignment between our approach to the video-sharing platform (VSP) regime and our approach to the Codes, noting that it was not apparent how we had incorporated learnings from the VSP regime.[620]

A1.15.10 Some stakeholders mentioned that our proposals for terms and statements should be in alignment with the Digital Services Act (DSA) to minimise the regulatory compliance obligations on service providers.[621]

A1.15.11 Refuge commented that providing clear and accessible terms and statements is "the very minimum companies should be doing to ensure their users are aware of the safety measures on their platforms".[622] In its response to our May 2024 Consultation, civil society group NEXUS NI suggested that services go beyond text-based terms and statements, including audio and visual mediums to help engage young people.[623] In its response to our

[618] Federation of Small Businesses response to November 2023 Consultation, p.4. We note that Federation of Small Businesses made a similar point in May 2024 Consultation, p.7.
[619] 5Rights Foundation response to November 2023 Consultation, p.25.
[620] Snap response to November 2023 Consultation, p.17.
[621] Mid Size Platform Group response to May 2024 Consultation, pp.11-12; Skyscanner response to November 2023 Illegal Harms Consultation, p.22; Snap response to November 2023 Consultation, p.17.
[622] Refuge response to November 2023 Consultation, p.18.
[623] NEXUS NI response to May 2024 Consultation, p.19.

May 2024 Consultation, the Canadian Centre for Child Protection (C3P) recommended specific standardised requirements for the measure on 'Clarity and accessibility of terms and statements' to ensure the clarity and accessibility of services' terms and statements.[624]

Our response

A1.15.12 In response to the Federation of Small Businesses' point, it would not be feasible for us to provide generators for the provisions required under the Act due to the range of service providers within the scope of this measure. There are also significant differences in the safety measures, systems, and processes used to protect users from illegal harms across various service providers.

A1.15.13 In response to 5Rights Foundations's point, where appropriate, we may, from time to time engage with services to check that their terms and statements are suitably aligned with the Illegal Content Judgements Guidance ('the ICJG').

A1.15.14 In response to Snap's point, while we have drawn upon the learnings from the VSP regime, online safety is a different regime and requires a specialised approach.

A1.15.15 In response to points relating to alignment with the DSA, our recommended measures for terms and statements codify the duties for service providers under the Online Safety Act (OSA) whilst providing adequate flexibility for service providers to design their terms and statements in a way that suits their service. Our measures will not prevent service providers from using the same terms and statements for both the OSA and the DSA.

A1.15.16 In response to points from Refuge, NEXUS NI and C3P, the measures we recommend are the minimum standard required for service providers to meet their duties in line with the Act. If we take the view that service providers are not meeting their duties, we will engage with them through the supervisory function or take enforcement action. It is also important to note that the Codes may be updated in future.

The role of service providers

A1.15.17 Glitch highlighted the importance of service providers consulting with advocacy groups and marginalised communities when developing terms and statements.[625]

A1.15.18 Bereaved Families for Online Safety raised concerns that service providers may use their terms of service to bypass their duties under the Act (for example, by stating that content does not violate community guidelines when it is reported by a user).[626]

A1.15.19 Big Brother Watch raised concerns about the terms of service models currently used by large social media services.[627] It argued that these models fail to adequately reflect human

[624] C3P response to May 2024 Consultation, p.26.
[625] Glitch response to November 2023 Consultation, p.10.
[626] Bereaved families for Online Safety response to November 2023 Consultation, p.2.
[627] Big Brother Watch response to November 2023 Consultation, pp.10-11. We note that Big Brother Watch made a similar point in May 2024 Consultation on Protecting Children from Harms Online, p.47.

rights principles, the rule of law, and democratic values, particularly when it comes to enforcement.

Our response

A1.15.20 In response to Glitch's point, we recognise the benefits of service providers taking such action. However, whilst they may choose to do so, it is not an approach that is required in order for service providers to comply with their duties under the Act.

A1.15.21 In response to the Bereaved Families for Online Safety's point, we cannot make judgements on individual pieces of content. Service providers are permitted to state that content does not breach their terms and statements if they judge this to be the case (provided that the content is not illegal). They are responsible for interpreting and applying their policies in individual cases, ensuring that this is done in a consistent way. Where there are reasonable grounds to infer that content is illegal under UK law, service providers cannot use their terms and statements to bypass this. There are also additional terms of service duties on Category 1 service providers to act in accordance with their terms and statements.[628] We will be consulting on these duties in 2025.

A1.15.22 In response to Big Brother Watch's point, Category 1 service providers will have additional duties to act in accordance with their provisions about taking down content, restricting access to content, or banning or suspending users to ensure greater transparency and accountability. These service providers will also have additional duties to protect journalistic content, content of democratic importance, and news publisher content.

A1.15.23 Under Section 22 of the Act, U2U service providers have a duty to consider freedom of expression and privacy.[629] There are also additional duties in this area for Category 1 service providers.[630] Where service providers fail to discharge their obligations under the Act, we will take enforcement action.

[628] Sections 71 and 72 of the Act.
[629] Section 22(2) and 22(3) of the Act.
[630] Section 22(4), 22(6) and 22(7) of the Act.

A1.16 Codes of Practice: User access

Stakeholder responses by theme

Measure on removing proscribed organisation accounts

A1.16.1 As mentioned in Volume 2 chapter 11, a majority of stakeholders were supportive of our proposed approach. One stakeholder said it did not agree with the measure but did not provide a reason as to why.[631]

A1.16.2 The Christchurch Call Advisory Network (CCAN) expressed concern about the destruction of important evidence for law enforcement as a result of this measure, suggesting that "removal and takedowns of certain types of content can result in harm, including destruction of evidence".[632] It further noted that "…such removals would hinder the UK's ability to uphold the Call [Christchurch Call to Action] commitment to: 'Ensure appropriate cooperation with and among law enforcement agencies for the purposes of investigating and prosecuting illegal online activity in regard to detected and/or removed terrorist and violent extremist content, in a manner consistent with rule of law and human rights protections'".[633]

A1.16.3 We also received a response from the Federation of Small Businesses that suggested "consideration should also be given to usernames which remain available following association with certain banned groups, and the reputational effect on other users if that username becomes available and they unknowingly take it up".[634]

Our response

A1.16.4 We recognise that recommending service providers remove accounts run by or on behalf of proscribed organisations carries a risk of inadvertently destroying potential evidence for law enforcement. However, service providers are in any case required to remove illegal content of which they become aware, and the Act does not give us powers to require providers to preserve evidence for criminal enforcement purposes.[635] As our powers to address illegal harms are therefore related specifically to the removal of illegal content, the decision to retain data from a removed account is a matter for the service provider. We do

[631] Centro de Estudios en Libertad de Expresión y Acceso a la Información (CELE) response to November 2023 Illegal Harms Consultation, p.13.
[632] Christchurch Call Advisory Network (CCAN) response to November 2023 Illegal Harms Consultation, p.3.
[633] CCAN response to November 2023 Consultation, p.3.
[634] Federation of Small Businesses response to November 2023 Illegal Harms Consultation, p.4.
[635] Section 10(3)(b) of the Act.

not consider that the risk that evidence may be lost is a reason not to recommend a measure which will reduce overall illegal activity and protect users.

A1.16.5 Our recommended measure does not include provisions relating to service providers seeking to prevent removed users from returning to the service, as this is an area in which we are still gathering evidence to ensure any recommendation is effective. However, this does not prohibit service providers from taking actions to prevent users from returning to their services. This stakeholder submission highlights the risk to user rights associated with preventing users from returning to a service and emphasise the importance of ensuring any such recommendation is effective and proportionate. Furthermore, the Online Safety Act ('the Act') does not require service providers to protect the reputation of their users. Therefore, we are not in a position to make formal recommendations pertaining to protecting the reputation of users.

Option explored on banning user accounts that spread CSAM

A1.16.6 In our November 2023 Consultation, we requested evidence to support our consideration of a possible future measure that would recommend blocking users that spread CSAM.[636]

Support

A1.16.7 Stakeholders were supportive of this possible measure.[637] In particular, some stakeholders, such as the Canadian Centre for Child Protection (C3P) and Marie Collins Foundation, indicated that any account that posts and shares CSAM should be permanently blocked from the service.[638] We also received feedback that demonstrated several service providers already ban accounts that spread CSAM.[639] Some examples of current banning practices included: (1) banning the username, email address, IP address (though

[636] A note on terminology: we have reflected that 'banning' is a more appropriate term to use in this context. As such, throughout this chapter, we have used the term 'banning' instead of 'block' or 'blocking', where appropriate. This avoids any potential confusion with the feature or function available to users who wish to block another user. We also note that the Act refers to 'banning' users (see e.g. sections 17(8), 18(12), and 71(3)(b)).

[637] Canadian Centre for Child Protection (C3P) response to November 2023 Illegal Harms Consultation, p.27; Centre for Competition Policy response to November 2023 Illegal Harms Consultation, p.17; GeoComply Solutions response to November 2023 Illegal Harms Consultation, pp.11-15; Marie Collins Foundation response to November 2023 Illegal Harms Consultation, p.18; Match Group response to November 2023 Illegal Harms Consultation, p.18; Scottish Government response to November 2023 Illegal Harms Consultation, p.11; Welsh Government response to November 2023 Illegal Harms Consultation, p.5.

[638] C3P response to November 2023 Consultation, p.27; Marie Collins Foundation response to November 2023 Consultation, p.18; Philippine Survivor Network response to November 2023 Illegal Harms Consultation, p.17; Segregated Payments Ltd response to November 2023 Illegal Harms Consultation, p.14; Snap response to November 2023 Illegal Harms Consultation, p.24.

[639] Name withheld 5 response to November 2023 Illegal Harms Consultation, pp.14-15; Google (confidential) response to November 2023 Illegal Harms Consultation, pp.67-68; [✂]; Microsoft response to November 2023 Illegal Harms Consultation, p.20; Name withheld 4 (confidential) response to November 2023 Illegal Harms Consultation, p.10; [✂]; UK Interactive Entertainment (Ukie) response to November 2023 Illegal Harms Consultation, pp.29-30; WeProtect Global Alliance response to November 2023 Illegal Harms Consultation, pp.23-24.

considered ineffective), and, in some cases, at the device level; or (2) suspending the offender's account by permanently preventing the device used to connect to the network.[640] In addition, one stakeholder provided recommendations on how to design a measure that bans accounts that share CSAM, including using [✂].[641] A different stakeholder expressed concern about banning IP addresses, stating they "do not identify individuals – they are frequently re-allocated and even when not reallocated, only tend to identify a household, which would cause people other than the targeted individual to be blocked".[642] One stakeholder said a strikes system would not be appropriate for CSAM given the severity of harm posed by this type of illegal content.[643]

A1.16.8 This evidence helps support future development of a measure to ban users that spread CSAM. This was reiterated by the National Society for the Prevention of Cruelty to Children (NSPCC).[644]

A1.16.9 However, one stakeholder did not agree with our proposed approach, but did not provide evidence to articulate why they were unsupportive.[645]

Our response

A1.16.10 We will consider these points when assessing the proportionality of any future measures.

Application of measure

A1.16.11 We received feedback that highlighted the need to mitigate the risk that accounts that are not sharing CSAM (i.e., non-violative) are not erroneously banned. This includes the risk of wrongfully identifying accounts as CSAM. The responses also referenced the length for which a user should be banned for sharing CSAM, and how the nature of the offence committed impacts this decision. These specific responses are:

- Meta suggested that user accounts should be permanently banned in the event of a single instance of malicious sharing of CSAM, and that the penalties for other CSAM violations should take into account the severity of harm and the number of strikes on the user's account.[646]
- Snap indicated that there are circumstances where a user is banned for sharing content that is flagged as CSAM, however, a permanent ban is not the most appropriate option. In these cases, context often matters and an understanding that CSAM can take various forms.[647]

---

[640] Ukie response to November 2023 Consultation, pp.29-30; WeProtect Global Alliance response to November 2023 Consultation, pp.23-24.
[641] [✂]
[642] Name withheld 2 response to November 2023 Illegal Harms Consultation, p.15.
[643] Institute for Strategic Dialogue (ISD) response to November 2023 Illegal Harms Consultation, p.12.
[644] National Society for the Prevention of Cruelty to Children (NSPCC) response to November 2023 Illegal Harms Consultation, pp.42-44.
[645] CELE response to November 2023 Consultation, p.13.
[646] Meta response to November 2023 Consultation, confidential annex, p.18
[647] Snap response to November 2023 Consultation, pp.24-25.

- Segregated Payments Ltd indicated that categorising the severity of an offence would be useful in cases of 'mistakes' or borderline content. There should also be an appeals process established for cases of wrongful identification of CSAM. However, it said that if verified CSAM is shared then the account should be permanently removed.[648]
- The Scottish Government recommended a risk-based approach to determining whether accounts should be banned that spread CSAM.[649]
- NSPCC recommended that the measure outline criteria for banning a user account that shares CSAM, which should consider the age of the offender, the nature of the CSAM, the intention, and whether it was a repeat offence.[650]
- The Philippine Survivor Network expressed the need for due process in banning users who share CSAM and that users should be banned permanently due to the likely risk of re-offending.[651]
- [✂] indicated that they have systems in place to account for potential false positives, including an appeal process for users to explain why the content does not violate the service's terms and conditions, or why they have a legitimate reason to possess it. Depending on the nature of the content, the potential restrictions placed on their account, and the information available to the service provider, the content may be re-reviewed and reinstated.[652]

A1.16.12 [✂][653] It was also suggested that services can deploy multi-factor authentication that uses more than one assessment service to mitigate this risk.[654] Other suggestions included that service providers can use various detection technologies and/or human review to mitigate content being erroneously classified as CSAM;[655] and service providers can ensure expedited appeals processes for wrongful classifications of illegal content.[656]

Our response

A1.16.13 We understand concerns regarding how a measure in this space could raise various difference risks. This is a point we will consider when assessing the proportionality of any future measures.

---

[648] Segregated Payments Ltd response to November 2023 Consultation, p.14.
[649] Scottish Government response to November 2023 Consultation, p.12.
[650] NPSCC response to November 2023 Consultation, p.43.
[651] Philippine Survivor Network response to November 2023 Consultation, p.17.
[652] Name withheld 5 response to November 2023 Consultation, p.14.
[653] [✂]
[654] Safe Space One response to November 2023 Illegal Harms Consultation, p.19.
[655] LinkedIn response to November 2023 Illegal Harms Consultation, p.17; WeProtect Global Alliance response to November 2023 Consultation, pp.24-25.
[656] ISD response to November 2023 Consultation, p.13; WeProtect Global Alliance response to November 2023 Consultation, pp.24-25.

Recidivism

A1.16.14 We received feedback from stakeholders saying that repeat offenders can often regain access to services after removal. We consider this to be an important risk to consider for a user access measure related to any illegal harm. Specifically, Microsoft noted "…we have implemented processes to combat instances of recidivism [✂] Tactics to address this problem require extensive human analysis and may compromise the privacy of legitimate users".[657]

A1.16.15 We also received some stakeholder feedback providing examples of how to prevent a user returning to a service which we will consider. Meta indicated that there are various mechanisms that can be used to prevent a user from returning to the service.[658]

A1.16.16 GeoComply Solutions, INVIVIA, recommended processes that can help identify suspicious location patterns and prevent users circumventing account bans.[659] These solutions included: device finger printing; authenticated multi-source geo location data; social authentication; biometric verification; and behavioural analysis and machine-learning technology.

Our response

A1.16.17 We recognise that preventing users from returning to a service is an important consideration for effectiveness. The prevention of removed users creating new accounts is a complex issue and we will consider this in our proportionality assessment of any measure suggested.

Risk

A1.16.18 In addition to listing potential options for blocking a user, Protection Group International also flagged that "consideration […] needs to be given to pushing offenders onto other platforms – could certain information be shared across platforms?".[660]

Our response

A1.16.19 This is a point we will consider when assessing the proportionality of any future measures.

Option explored on broad strike and blocking systems

A1.16.20 In our November 2023 Consultation, we asked for views on recommending a broad strike and blocking system as a measure for service providers. This section outlines the stakeholder feedback received pertaining to a strike and banning system against users where they are found to have posted or shared illegal content or committed or facilitated illegal behaviour.

---

[657] Microsoft response to November 2023 Consultation, p.20.
[658] Meta response to November 2023 Consultation, confidential annex, p.17.
[659] GeoComply Solutions response to November 2023 Consultation, pp.11-15; INVIVIA response to November 2023 Illegal Harms Consultation, p.27;
[660] Protection Group International response to November 2023 Illegal Harms Consultation, p.12.

Extending measure to other illegal harms

A1.16.21 We received feedback from several stakeholders expressing support for broadening the scope of the measure to address other illegal harms. Some stakeholders - the Local Government Association,[661] the Cyber Helpline,[662] and [✂][663] - argued for a more general expansion of the types of illegal harm captured by this measure. Other stakeholders recommended specific illegal harms to which a blocking and strikes system should be applied:

- **Fraud**: [✂],[664] Cifas,[665] and UK Finance[666] suggested the measure should be extended to cover fraud and offences under the Proceeds of Crime Act 2002 (POCA) and Financial Services Markets Act 2000 (FSMA). [✂].[667] Booking.com indicated that they already suspend or terminate accounts participating in fraudulent activity.[668]
- **Firearms**: The [✂] suggested the measure should be expanded to those supplying firearms.[669]
- **Organised Crime**: C3P proposed that the measure should be expanded to organised criminal activities.[670]
- **Hate**: The Board of Deputies of British Jews argued the measure should be applied to user accounts that espouse antisemitic views on services in scope. They also suggested that investigatory mechanisms be used to identify anonymous or dummy accounts operated by previously banned users.[671]
- Another shared their service blocked for posts that were not aligned with the purpose of the forum, not just illegal or harmful content.[672]

Our response

A1.16.22 As indicated in the November 2023 Consultation, we need more evidence to be able to recommend a proportionate blocking measure that applies to more types of illegal harms, which includes appropriate safeguards for user rights.

Strikes

A1.16.23 We also received responses from three stakeholders – the Institute for Strategic Dialogue ('ISD'), the Local Government Association, and the Online Safety Act Network ('OSA Network') – on the viability of account strikes. The ISD stated that the enforcement and

---

[661] Local Government Association response to November 2023 Consultation, p.14.
[662] The Cyber Helpline response to November 2023 Illegal Harms Consultation, p.19.
[663] [✂]
[664] [✂]
[665] Cifas response to November 2023 Illegal Harms Consultation, p.18.
[666] UK Finance response to November 2023 Illegal Harms Consultation, pp.2, 11.
[667] [✂]
[668] Booking.com response to November 2023 Illegal Harms Consultation, p.9.
[669] [✂]
[670] C3P response to November 2023 Consultation, pp.26-27.
[671] The Board of Deputies of British Jews response to November 2023 Illegal Harms Consultation, pp.2-3.
[672] Bolton, C. response to November 2023 Illegal Harms Consultation, p.10.

implementation of any account blocking policies are not visible to external organisations, and stated that "account strikes are not an appropriate response to the sharing of highly harmful illegal content, such as terrorist content or CSAM".[673] The Local Government Association suggested that it was reasonable to remove the accounts of those "who frequently or persistently undertake dissemination of illegal content, harass or intimidate via a service, or who persistently establish accounts to disseminate false communications".[674] Lastly, OSA Network stated there is "no consideration of the rights protected through blocking, or the value ascribed to the speech in the blocked account as regards to both the speakers' rights and those receiving the information". It also noted that measures for CSAM blocking are not included in the draft Illegal Content Codes of Practice despite the impact on children and likely interference with children's Article 8 and 3 rights and also possibly Articles 2 and 4.[675]

Our response

A1.16.24 As indicated in the November 2023 Consultation, we need more evidence to be able to recommend a proportionate blocking measure that applies to more types of illegal harms, which includes appropriate safeguards for user rights.

Banning

A1.16.25 The Domestic Abuse Commissioner provided feedback concerning the measure's potential effect on the rights of victims. It provided a cross-cutting critique of the approach in the Codes arguing that positive effects for victims of online harms should be prioritised over considerations of users' rights to privacy or freedom of expression. Specifically, it stated that "when balancing freedom of expression rights with recommending measures for strikes or blockings, Ofcom considers users mainly as 'speakers' and focuses on their rights in terms of freedom of expression. However, when thinking about the Human Rights Act, Ofcom fails to consider on balance the rights of users in terms of protection through blocking: 'although blocking and strikes may be a way of tackling illegal content, there are also concerns about the use of these systems on lawful speech.' It [is] important to recognise that, while the 'takedown' approach that Ofcom has focused on will indeed assist individuals on case-by-case bases to take down content once it has already been posted, a better approach would be to think about the system in place. There can be better design choices being made by services and encouraged by Ofcom, and recommendations should be based more on design choices which perpetuate misogynistic/harmful behaviours with clear intent. The consultation should be focused on upstream safety by design, not reactive takedown measures."[676] This response did not suggest a need to change the proposed measure to remove the accounts of proscribed organisations.

---

[673] ISD response to November 2023 Consultation, p.12.
[674] Local Government Association response to November 2023 Consultation, p.14.
[675] Online Safety Network Act Network (OSA Network) response to November 2023 Illegal Harms Consultation, annex C, p.3.
[676] Domestic Abuse Commissioner response to November 2023 Illegal Harms Consultation, p.7.

A1.16.26 Similarly, The Community Security Trust and Antisemitism Policy Trust said that "Ofcom's approach to blocking users also considers limitations to these users' freedom of expression and freedom of association, but not the freedom of expression of their intended victims which could be violated by users' illegal speech."[677]

Our response

A1.16.27 We have addressed similar feedback received on our overall approach in the 'Our approach to developing Codes' chapter.

Consideration of identity verification

A1.16.28 In the November 2023 Consultation, we set out the reasons we were proposing not to recommend an identity verification measure. In their responses, several stakeholders commented on identity verification, albeit not necessarily responding specifically to our proposal. We also received general feedback on identity verification in response to other questions in our consultation. We have addressed all of the feedback related to identity verification in this section.

Requests for further measures

A1.16.29 Several respondents argued for a measure that enables users to voluntarily verify their identity. The Community Security Trust and Antisemitism Policy Trust agreed that having an option to verify users' identity would allow those with protected characteristics to express themselves more freely.[678] Several stakeholders suggested measures that paired optional verification schemes with the ability for users to control experiences on services. Clean up the Internet submitted a proposal that certain platforms offer their users a choice of verifying, label accounts that have verified, and give users enhanced controls to manage their interaction with non-verified accounts.[679] [✂] said voluntary verification should be in place specifically for social media and dating services, noting the level of harm caused by romance fraud; OneID suggested such a measure includes an ability for verified users to choose not to see content of non-verified users.[680] Yoti recommended the use of optional user verification as a common-sense measure to reduce and prevent illegal harms across a wide range of platforms. It also expressed willingness to engage with Ofcom as it explores this approach, and options for individuals to filter out non-verified users.[681] LoveSaid supported more services adopting verification schemes.[682]

A1.16.30 There were other responses that suggested we include other types of identity verification measures. The Philippine Survivor Network urged that there should be mandatory identity

---

[677] Community Security Trust and Antisemitism Policy Trust response to November 2023 Illegal Harms Consultation, p.11.
[678] Community Security Trust and Antisemitism Policy Trust response to November 2023 Consultation, p.14.
[679] Clean up the Internet – Proposal for a measure requiring platforms to offer their users options to verify their identity, response to November 2023 Illegal Harms Consultation, p.2.
[680] [✂]; OneID response to November 2023 Illegal Harms Consultation, p.3-4.
[681] Yoti (confidential) response to November 2023 Illegal Harms Consultation, pp.17-18.
[682] LoveSaid response to November 2023 Illegal Harms Consultation, p.15.

verification for any user account across all services to mitigate newly produced CSAM.[683] Yoti and Innovate Finance suggested that we set minimum standards for identity verification methods.[684]

A1.16.31 Innovate Finance, UK Finance, OneID, [✂], [✂], and [✂] advocated for verification of sellers in online marketplaces, citing it as a necessary intervention against purchase scams occurring on such services.[685] Innovate Finance specifically provided evidence on the risks associated with certain services. UK Finance also suggested mandatory verification for those using social media and dating platforms to mitigate harm caused by perpetrators benefitting from anonymity and pseudonymity.

A1.16.32 Another stakeholder encouraged cross-platform collaboration to verify user accounts.[686]

Our response

A1.16.33 We will be taking a holistic view of identity verification as we progress our work on categorised services. We will take these responses into account as we do that work.

A1.16.34 We are exploring the implementation of optional identity verification for services as part of our Phase 3 work on the user identity verification guidance, which the Act requires us to publish.[687] This involves considering many of the aspects raised by stakeholders here. Following this, we will be able to consider how we may incorporate identity verification into other measures in the future.

Anonymity

A1.16.35 We also received responses highlighting the risk of anonymity, and the potential of identity verification to mitigate this risk. Another stakeholder indicated the ability to make user profiles anonymous may embolden users to engage in stalking behaviours without fear of repercussions.[688] They further expressed concerns that our consideration that identity verification requirements would impede too heavily on user's freedom of speech deprioritises the safety of users and the detection and prevention of crimes.[689] Geocomply Solutions expressed support of our assessment of the risks associated to pseudonymity and anonymity. It also supported the measure's efforts to strengthen identity verification protocols and mechanisms used to prevent account recidivism, which is required to effectively prevent CSAM-related harms.[690]

---

[683] Philippine Survivor Network response to November 2023 Consultation, pp.10-11.
[684] Innovate Finance response to November 2023 Illegal Harms Consultation, pp.11-12; Yoti response to November 2023 Illegal Harms Consultation, p.17.
[685] [✂]; Innovate Finance response to November 2023 Consultation, pp.11-12, 16; [✂] [✂]; OneID response to November 2023 Consultation, p.4; UK Finance response to November 2023 Consultation, p.13.
[686] Lovesaid response to November 2023 Consultation, p.15.
[687] More information about categorised services and Phase 3 can be found at Ofcom's approach to implementing the Online Safety Act.
[688] Suzy Lamplugh Trust response to November 2023 Illegal Harms Consultation, pp.9-10.
[689] Suzy Lamplugh Trust response to November 2023 Consultation, p.10.
[690] Geocomply Solutions response to November 2023 Consultation, pp.1-3.

A1.16.36 Conversely, Reddit, Mid Size Platform Group, Nexus, and the Centre for Competition Policy highlighted the importance users attach to online anonymity and/or pseudonymity.[691]

A1.16.37 In a similar vein, one stakeholder stated that requiring individuals to provide identification to services risks the loss of user privacy and could result in users being blocked across all services as a result of being blocked from one service.[692]

Our response

A1.16.38 In the November 2023 Consultation, we communicated that there are many legitimate reasons for users to have multiple accounts or to not verify their identity, and that often anonymity or pseudonymity is important to enable certain users to exercise their rights to freedom of expression and association.

A1.16.39 We recognise the risks posed by user anonymity, as outlined in our Register of Risks ('Register') which details several harms that can be facilitated by levels of concealment of a user's identity, including anonymity and pseudonymity. We also recognise that the impact on user rights of any measure which recommends mandatory identity verification would need to be carefully assessed to ensure it is proportionate.

A1.16.40 We will consider these points in our assessments of proportionality for future measures.

Age assurance

A1.16.41 In our November 2023 Consultation, we set out our strategy and future consultations on age assurance. In response, we received responses from stakeholders regarding the use of age assurance requirements for specific types of illegal harms. These included:

- One stakeholder encouraged the use of age verification for protecting children online.[693]
- Other stakeholders proposed the use of age assurance to combat harms including, self-generated intimate imagery and detect unknown CSAM.[694] Similarly, one stakeholder argued that services advertising sex work should be behind a robust age verification system.[695]

[691] Centre for Competition Policy response to November 2023 Consultation, p.17; Mid Size Platform Group response to November 2023 Illegal Harms Consultation, p.11; Nexus response to November 2023 Illegal Harms Consultation, p.18; Reddit response to November 2023 Illegal Harms Consultation, p.22.
[692] Name withheld 3 response to November 2023 Illegal Harms Consultation, p.18.
[693] The Independent Inquiry into Child Sexual Abuse response to November 2023 Illegal Harms Consultation, p.4.
[694] Age Verification Providers Association's response to November 2023 Illegal Harms consultation, pp.2-3; Online Safety Technology Industry Association (OSTIA) response to November 2023 Illegal Harms Consultation, p.15; VerifyMy response to November 2023 Illegal Harms Consultation, p.8.
[695] Nordic Model Now response to November 2023 Illegal Harms Consultation, p.18.

- One stakeholder expressed concern over the bias gaps that exist with age estimation tools, which may impact the accuracy of the tools and leave users with false security.[696]
- One stakeholder commented on the language we use to describe specific age assurance elements.[697] In response, we reviewed any reference to age assurance to ensure it aligns with the definitions outlined the Act.

Our response

A1.16.42 We consulted on the use of highly effective age assurance to protect children from harmful content in May 2024. We will consult on how age assurance can be used to protect children from grooming in Spring 2025. We will consider requests made for further age assurance measures when developing future iterations of the Codes.

[696] C3P response to November 2023 Consultation, pp.32, 34.
[697] Yoti response to November 2023 Consultation, pp.17-18.

A1.17 Codes of Practice: User controls

Stakeholder responses by theme

User account blocking and disabling comments

Retaining information about blocked accounts and sharing details of blocked users across services owned by the same provider

A1.17.1 The Canadian Centre for Child Protection (C3P) expressed concern that we were not recommending that services should retain information about blocked accounts to allow them to identify perpetrator accounts more quickly. It also suggested that services owned by the same provider should share blocked account details with one another.[698]

Our response

A1.17.2 We acknowledge the issue of perpetrators persistently targeting individuals online and have recommended that services implement a feature that allows users to block all unconnected users. At this point, we consider that this goes far enough to protect users who are persistently targeted by the same perpetrator(s). The effectiveness of tracking blocked user details and sharing them across services is unclear, given that some blocked users create accounts with false details to try and circumvent individual blocks. This would also raise other considerations, such as user privacy, data protection, and the additional burden that implementing such policies would place on services.

Geo-fencing and tracking IP addresses

A1.17.3 C3P said that there should be an opportunity to know where a user is using the same IP address to repeatedly commit harm, and that users should be able to limit who can contact them outside of a certain geographical radius, such as their own city.[699]

Our response

A1.17.4 We have not considered IP address tracking or geo-fencing for these measures. It is unclear how effective this would be in preventing harm, as virtual private networks (VPNs), proxies and other privacy tools can mask a user's real IP address and location. This would also raise other considerations, such as the effect it would have on innocent persons using shared IP addresses.

[698] The Canadian Centre for Child Protection (C3P) response to November 2023 Illegal Harms Consultation, p.25.
[699] C3P response to November 2023 Consultation, p.26.

Terror offences and fraud

A1.17.5 The Independent Reviewer of Terrorism Legislation argued that the measures should extend to terrorism content.[700]

A1.17.6 Cifas said that the blocking measure should be extended to services with a medium or high risk of fraud, as direct messaging is a way for perpetrators of fraud to contact their victims.[701]

Our response

A1.17.7 All illegal harm content types were considered during the development of these measures. We do not have sufficient evidence of how blocking or muting and comment controls would relate to effective prevention of specific terrorism or fraud harms. We will continue to review the evidence to assess whether we should expand these measures to other kinds of illegal harms in future iterations of the Codes.

Virtual reality

A1.17.8 SPRITE+ said that we should consider the addition of virtual reality platform-appropriate user controls.[702]

Our response

A1.17.9 At this stage, we have not considered requiring specific user controls for different types of services. We consider that our recommended measures are appropriate for all service types that fall within scope. We will continue to review the impact of our measures and may consider specific recommended user controls for different service types if we find any evidence that this could be beneficial and proportionate.

Bulk tools

A1.17.10 The Institute for Strategic Dialogue argued for the provision of bulk reporting, blocking, and muting tools for victims of online harassment or abuse, especially for high-profile, vulnerable, or marginalised users.[703]

Our response

A1.17.11 The ability to block all unconnected users on a service goes a long way towards providing this sort of functionality. We will monitor the impact of this measure and will consider iterations if and when we encounter any evidence to suggest this is necessary.

A1.17.12 Regarding bulk reporting, measures ICU D2 and ICS D2 recommend that service providers allow users and affected or interested persons to provide relevant information or supporting material when submitting complaints. Our measures allow service providers

---

[700] The Independent Reviewer of Terrorism Legislation response to November 2023 Illegal Harms Consultation, p.7.

[701] Cifas response to November 2023 Illegal Harms Consultation, p.17.

[702] SPRITE+ (University of Glasgow) response to November 2023 Illegal Harms Consultation, p.16.

[703] Institute for Strategic Dialogue response to November 2023 Illegal Harms Consultation, p.12.

flexibility over how they achieve this. In some cases, users may be able to use this functionality to report several pieces of content at once (depending on how services design their complaints functions), which may offer similar functionality to bulk reporting.

Government accounts blocking citizens

A1.17.13 Centro de Estudios en Libertad de Expresión y Acceso a la Información (CELE) said that allowing government accounts and public officer accounts to block citizens could prevent blocked citizens accessing information.[704]

Our response

A1.17.14 While we appreciate these concerns, we maintain that all users should have equal access to blocking functionality. This includes those who hold high profile positions in public life, who we know can be subjected to harmful behaviour on U2U services.[705]

Offering users the option to report accounts when blocking them

A1.17.15 The British and Irish Law, Education and Technology Association (BILETA) said that users should be given the option to report the accounts they are blocking and muting at the point they are blocking or muting them.[706]

Our response

A1.17.16 Our reporting and complaints measures refer to reporting content, rather than user profiles. We have not considered introducing a requirement for services to enable users to report profiles, but we will be iterating our approach to the Codes quickly and will review this if we receive any evidence to suggest a change would be both proportionate and beneficial for users.

A1.17.17 Regarding children, our supportive messaging measure (ICU F2) requires services to inform child users of the options available to them to increase their safety when they take action against another account (such as blocking, muting, or reporting a user's conduct). While not specifically requiring services to give users the chance to report at the point of blocking, we consider this measure will result in similar outcomes for users. It will enable children to make informed choices about their safety in their online experiences. We will continue to review the evidence in this area.

---

[704] Centro de Estudios en Libertad de Expresion y Acceso a la Informacion (CELE) response to November 2023 Illegal Harms Consultation, p.12.
[705] As set out in the 'Benefits and effectiveness' section relating to the blocking, muting, and disabling comments measures.
[706] The British and Irish Law, Education and Technology Association (BILETA) response to November 2023 Illegal Harms Consultation, p.16.

Educational resources for users

A1.17.18 SPRITE+ said the measures could be strengthened by providing users with information relating to how recommender systems promote content to them, along with more general educational resources on how to stay safe online.[707]

Our response

A1.17.19 We recognise that educational resources may help keep users safe online. We also recognise that providing users with information about how recommender systems curate content feeds may promote understanding amongst users. For example, accessible explanations of what affects content ranking may prompt users to engage more critically with the content that is recommended to them. However, we have not considered requiring services to implement such measures at this time because we have decided to introduce our first set of Codes quickly to ensure immediate impact for users. The introduction of such measures would require further evidence of how such information might affect user behaviour, particularly in relation to recommender systems. Separately, there are wider concerns about such explanations containing information that could be used by perpetrators. We will continue to monitor this area and will consider measures in this space should the evidence base suggest that they would be proportionate.

Recommending these measures in the Illegal Content Codes of Practice

A1.17.20 Google and techUK argued that measures IHUJ1 and IHUJ2 should be recommended as part of Ofcom's work on the user empowerment duties in Phase 3 of implementation. Google said that, as the user empowerment duties provide detail on the types of controls platforms should offer to users, and that these duties only apply to Category 1 services, Ofcom has gone beyond the scope of the Act by recommending these measures in the Illegal Content Codes.[708] TechUK argued against recommending the measures as part of these Codes because "the risk of misaligned obligation could create difficulties for platforms to understand and implement in a cohesive way".

Our response

A1.17.21 Regarding Google's response, the Act sets out the that the illegal content safety duties apply across all areas of a service, and expressly states in section 10(4)(f) that the duties require providers to take measures in relation to "functionalities allowing users to control the content they encounter". We consider that these measures fall within this area, and the consultation set out evidence of how they can reduce the risk of a variety of illegal harms. Recommending these measures in relation to the illegal content safety duties is therefore in scope of Ofcom's powers in the Act. In any case, the user empowerment duties will apply only to Category 1 services and will not apply to illegal content but rather

[707] SPRITE+ (York St John University) response to November 2023 Consultation, p.17.
[708] Google response to November 2023 Illegal Harms Consultation, p.59; techUK response to November 2023 Illegal Harms Consultation, p.28.

to certain specified categories of legal content.[709] As such, services which do not fall within scope of the user empowerment Code of Practice but would fall within scope of these measures could leave their users exposed to the risk of the harms to which these measures apply.

A1.17.22 We recognise that misalignment between measures which recommend similar functionality for different duties could unnecessarily increase the burden on service providers, as highlighted by techUK. As our recommendations must be proportionate, we seek to reduce that burden as far as possible. This includes ensuring that similar measures which are recommended in relation to different duties are aligned as far as possible. We intend to take the same approach for the user empowerment duties, and stakeholders will have the chance to comment on a draft user empowerment Code of Practice through consultation.

Notable user and monetised labelling schemes:

Risks to children

A1.17.23 5Rights Foundation noted that children struggle to distinguish between reliable and unreliable information online, which makes them susceptible to harm.[710] The Age Verification Providers Association suggested the measure should apply to child users by default.[711]

Our response

A1.17.24 We agree that some users face increased risk from harm and considered this when developing the measure. We do not have evidence that suggests relevant schemes should be designed in a particular way to protect children from fraud or foreign interference. Our expectation that providers should ensure user facing information is "written to a reading age comprehensible for the youngest individual permitted to use the service without the consent of a parent or guardian" should enable children accessing the services to understand such schemes.

Verifying medical practitioners

A1.17.25 Safe Space One suggested that the credentials of medical practitioners should be verified to reflect specific risks in that sector.[712]

Our response

A1.17.26 Providers may choose to label profiles of medical practitioners on the basis of their role or expertise depending on the intention of their scheme. However, we do not have sufficient evidence that this issue should be specifically addressed by this measure. The measure may

[709] See section 16(2) to (5) of the Act.
[710] 5Rights Foundation response to November 2023 Illegal Harms Consultation, pp.31-32.
[711] Age Verification Providers Association response to November 2023 Illegal Harms Consultation, p.3.
[712] Safe Space One response to November 2023 Illegal Harms Consultation, p.17.

help mitigate these risks should providers choose to have a policy on verifying the credentials of medical practitioners.

Evaluation of metrics

A1.17.27 Glitch, the Online Safety Act Network (OSA Network), and INVIVIA gave general feedback on the importance of service providers developing evaluation metrics to test the effectiveness of the measures they have applied.[713]

Our response

A1.17.28 We consider that our inclusion of expectations on providers to review their schemes, taking into account external feedback, sets proportionate expectations for this measure. We also expect our governance and accountability measures, which includes recommending service providers carry out yearly annual review of risk management activities, will go some way in doing this. See Volume 1: chapter 5: 'Governance and accountability' for more details on our package of measures.

Disagreement with need for measure

A1.17.29 Name Withheld 3 stated that issues with relevant schemes can be resolved through external scrutiny and argued that regulatory intervention was not necessary.[714]

Our response

A1.17.30 We maintain that the existing evidence of harm that has not been resolved by external scrutiny means intervention in the form of this measure is justified and proportionate.

Consideration of advertiser and provider content

A1.17.31 One respondent said it was unclear how our recommendations address fraudulent advertising and asked whether provider content, such as blue ticks, could be changed to mitigate risks.[715]

Our response

A1.17.32 This measure has not been designed to apply to advertising. It aims to help users assess the risk posed by content that has been posted by users participating in a relevant scheme.

A1.17.33 We do not think it is necessary to set expectations about the design of the labels providers may use to indicate that a user is participating in a relevant scheme. There will be a future Ofcom consultation on a distinct code of practice on requirements in relation to fraudulent

[713] Glitch response to November 2023 Illegal Harms Consultation, p.11; INVIVIA response to November 2023 Illegal Harms Consultation, p.24; Online Safety Act Network (OSA Network) response to November 2023 Illegal Harms Consultation, p.19.
[714] Name Withheld 3 response to November 2023 Illegal Harms Consultation, p.18.
[715] [✂].

advertising.[716] The measure instead focusses on helping users to understand how a provider's schemes operate and why a user profile has been labelled.

Harm to minority groups

A1.17.34 BILETA suggested there could be risks associated with minority groups and their anonymity on certain services, noting that labelling may expose these vulnerable individuals or groups to abuse.[717]

Our response

A1.17.35 At present, we have not identified a need to address the risk of minority groups being labelled against their will by providers under relevant schemes. If evidence of this emerges in the future, we will take it into account as appropriate.

Privacy implications

A1.17.36 Mid Size Platform Group suggested that voluntary verification schemes could create privacy concerns.[718]

Our response

A1.17.37 We agree with Mid Size Platform Group that voluntary labelling schemes could give rise to privacy implications. As such schemes are voluntary, we consider the impacts of our measure to be proportionate, as set out in the privacy section of our 'Rights impact'.

[716] More information on Ofcom's future work can be found in our Approach to Implementing the Online Safety Act.
[717] BILETA response to November 2023 Consultation, p.17.
[718] Mid Size Platform Group response to November 2023 Illegal Harms Consultation, p.11.

A1.18 Illegal Content Judgements Guidance

Introduction

A18.1 In this annex we outline other decisions on the Illegal Content Judgements Guidance (ICJG) which we have not covered in Volume 3, Chapter 2. These decisions largely relate to stakeholder responses to our November 2023 Illegal Harms Consultation and our August 2024 Further Consultation on Torture and Animal Cruelty.

Cross-cutting responses and decisions

Language used for trusted flaggers

A18.2 The Canadian Centre for the Protection of Children (C3P) recommended considering language like 'reporting entity or body' in addition to trusted flagger, or clarifying the definition of trusted flagger so it is not based on company discretion.[719]

Our response

A18.3 We accept this point and have amended the ICJG accordingly.

Process for service providers to alert Ofcom about availability of information

A18.4 In relation to our assessment of what information is reasonably available and relevant to regulated service providers when making illegal content judgements, LinkedIn encouraged Ofcom to ensure there is a process in place for regulated service providers to highlight issues they encounter with the availability of this information in practice and areas where further clarity is required.[720]

Our response

A18.5 We are committed to engaging with platforms on these and other issues through our Supervision regime, as well as through appropriate engagement with smaller platforms.

[719] Canadian Centre for Child Protection (C3P) response to November 2023 Illegal Harms Consultation, p. 34.
[720] LinkedIn response to November 2023 Illegal Harms Consultation, p. 20.

Presumption that service providers act in good faith

A18.6 One individual questioned the presumption they felt Ofcom had made that providers typically act in good faith.[721]

Our response

A18.7 We note the stakeholder's point but did not make any presumptions one way or the other about providers in preparing our guidance, and do not think it would make a practical difference to our guidance if we did.

Data protection: reference to Article 10 of GDPR

A18.8 The Information Commissioner's Office recommended that paragraph (A)1.70 of the introduction should specifically reference Article 10 of UK GDPR.[722]

Our response

A18.9 We have decided to add this reference to the ICJG. We have also expanded our drafting on privacy in the ICJG's Introduction chapter, at paragraphs 1.6 to 1.7.

Offence areas at particular risk of malicious reporting

A18.10 We are aware that some offence areas are particularly at risk of malicious reporting from bad actors or perpetrators, including in the case of controlling and coercive behaviour.

A18.11 As such, we have decided to add drafting to the section on malicious reporting at paragraph 1.769 to 1.70 in the ICJG's Introduction chapter to alert service providers to this risk.

[✂]

A18.12 [✂][723]

Our response

A18.13 [✂]

Reference to general defences

A18.14 In its response, the British and Irish Law, Education and Technology Association (BILETA) questioned our approach to general defences, arguing that they had been "dismissed quite lightly."[724]

---

[721] Are, C. response to November 2023 Illegal Harms Consultation, pp. 22-23.
[722] Information Commissioners Office (ICO) response to November 2023 Illegal Harms Consultation, p. 25.
[723] [✂]
[724] SPRITE+ (University of Sheffield) response to November 2023 Consultation, p. 23. British and Irish Law, Education and Technology Association (BILETA) response to November 2023 Illegal Harms Consultation, p. 22.

Our response

A18.15 We believe we have dealt with general defences in an appropriate way in our guidance, in line with what is reasonable to expect from services making content judgments at scale with access to limited information. We believe that general defences are unlikely to be relevant to a service provider's illegal content judgments as it is difficult to imagine circumstances in which service providers would have reasonable grounds to infer that they arise.

Definition of illegal content

A18.16 We note that illegal content is a new and complex concept, and that this may require further definition.

A18.17 We have therefore decided to expand Box 1 in the Introduction to give a more complete definition of illegal content.

Reference to Children's Harms Guidance

A18.18 Since publishing our consultation on the ICJG, Ofcom has also launched a consultation on Children's Harms Guidance which will be published in 2025.

A18.19 We have therefore decided to add a placeholder reference to all relevant chapters signalling where hyperlinks to the published Children's Harms Guidance will be found when it has been published in its final form.

Terrorism

Statutory definition of terrorism and risk of overly cautious or inappropriate moderation

A18.20 The Cyber Threats Research Centre, Swansea University argued that the UK statutory definition of terrorism and the offences in schedule 5 are too broad. This, it argued, may lead to "significant discretion vested in decision makers such as human moderators or automated tools," which may lead to an overly cautious or inappropriate application of the ICJG with negative impacts on freedom of expression.[725]

Our response

A18.21 Ofcom is only able to apply the law as it exists, and the fact that these offences are priority offences under the Act means that Parliament expressly turned its mind to them when it legislated. Our ICJG reflects the law as it currently stands. We have therefore made no change to the ICJG as a result of this response.

[725] Cyber Threats Research Centre at Swansea University response to November 2023 Illegal Harms Consultation, p. 16.

Language used in section on noxious substances

A18.22 One stakeholder, [✂] made a number of small points regarding noxious substances. This included suggested amended wording.[726]

Our response

A18.23 We have decided to action most of the points raised by the stakeholder, including amending the phrase "inferring the satisfaction of these criteria" to be clearer. However, we did not add the additional wording to state that, while "jokes, fantasies and fiction are unlikely to meet this test" they may meet the elements of another offence, as we believe that this is unlikely in most instances and to suggest as much would pose freedom of expression concerns.

Section 3 of the Terrorism Act: notices from constables

A18.24 The Independent Reviewer of Terrorism Legislation argued that Ofcom should not refer to section 3 of the Terrorism Act 2006 as it has never been used and "[providers] may be induced to believe that they can rely upon constables to do their assessments for them by serving a notice."[727]

Our response

A18.25 We acknowledge this point but have decided to retain reference to section 3 in our text. Although this facility has never been used, it may be in future and we want to ensure that service providers to understand their requirements under section 3 if they receive a notice. We have decided to add additional text to make it clear that "Providers should not rely solely on such notices to prompt action on content."

Wording regarding mental element of preparation of terrorist acts offence

A18.26 One stakeholder, [✂], suggested changing "high state of mind requirement" at A2.76 (now 2.105) to "significant mental element."[728]

Our response

A18.27 We believe "state of mind" is more comprehensible to a general audience than "mental element." However, we have deleted the adjective.

---

[726] [✂]
[727] Independent Reviewer of Terrorism Legislation response to November 2023 Illegal Harms Consultation, p. 7.
[728] [✂]

Expansion of section on terrorist financing

A18.28 One stakeholder, [✂], argued for an expansion of the section on terrorist financing, providing some suggested wording.[729]

Our response

A18.29 We have decided to amend our drafting to summarise the terrorist finance offences which are priority offences. These are now set out in bullet point form. The suggested text has been amended to be more understandable for a general audience.

Political speech and proscribed organisation offences

A18.30 X highlighted a perceived mismatch between paragraph 6B.4 of Volume 5 and Annex 10 and sought further clarity about users' rights relating to political discourse that may be shared with a proscribed organisation. In particular, X stated that "In practice, people may express support for groups labelled as terrorists under national laws due to political ideology rather than endorsing violence. Para 6B. 4 of Volume 5 read in isolation suggests that a user expressing an opinion or belief that is shared with a proscribed organisation is illegal and warrants the removal of content. However, Ofcom's Annex 10 clarifies that users must invite others to support a proscribed organisation to cross the criminal threshold. X seeks further criteria to distinguish when users' political discourse about or relating the opinions or beliefs of proscribed organisations crosses from protected freedom of expression into harmful content warranting removal."[730]

Our response

A18.31 We can clarify that political speech is not protected if it also amounts to a proscribed organisation offence. The ICJG sets out that content that can be reasonably inferred to amount to a proscribed organisation offence must be removed, and this applies regardless of whether it is considered political speech by the poster.

Technical corrections

A18.32 The Independent Reviewer of Terrorism Legislation noted two corrections regarding the mental elements of section 58 and section 12(1A)(B) of the Terrorism Act 2000. [731] He identified an error in our use of an authority, (Attorney General's Reference (No 4 of 2002) [2003] EWCA Crim 762).[732] Another stakeholder, [✂],suggested a change to the Legal Annex drafting on section 13(1A) of the Terrorism Act 2000.[733]

---

[729] [✂]
[730] X response to November 2023 Illegal Harms Consultation, p. 4.
[731] The Independent Reviewer of Terrorism Legislation response to November 2023 Consultation, p. 7.
[732] The Independent Reviewer of Terrorism Legislation response to November 2023 Consultation, p. 7.
[733] [✂]

A18.33 [✂] flagged that the preparation of terrorist acts offence is not "assisting others to commit such acts", which is the offence under S5(1)(b), but "engaging in preparation to assist others to commit such acts."[734]

Our response

A18.34 We have accepted all of these points and amended our texts accordingly.

Proscribed organisation offences and 'hyperlocal' context / dark humour

A18.35 Our draft ICJG stated that, when considering proscribed organisations offences, "cultural context will be particularly important and service [providers] should always take account of (hyper)local considerations when making judgements of this kind. This includes the different forms of humour deployed by online subcultures." We were concerned that this was not sufficiently exact and did not make it clear that the issue at hand is that some online subcultures employ dark or 'edgy' humour which may not appear to be a joke outside that subculture.

A18.36 We have decided to alter our drafting so that it now reads: "Cultural context will be particularly important and *providers should be mindful of the use of dark or 'edgy' humour in particular online subcultures*" (emphasis denotes change).

Threats, abuse and harassment (including hate)

Proposed ban on conversion therapy

A18.37 Humanists UK highlighted that the Government plans to ban conversion practices (practices intended to change someone's sexual orientation or gender identity), which may have implications for our guidance.[735]

Our response

A18.38 In the event that any changes to the law make it appropriate to review the ICJG, we will do so.

Definition of religious belief

A18.39 In its response, Humanists UK suggested that we had omitted non-belief from our definition of religious hatred.[736]

[734] [✂]
[735] Humanists UK response to November 2023 Illegal Harms Consultation, pp. 14-15.
[736] Humanists UK response to November 2023 Consultation, p. 15.

Our response

A18.40 Paragraph 3.52 (A3.46 in the draft ICJG) states that 'Religious hatred includes hatred against people defined by their religious belief, and hatred against people without religious belief (e.g. atheists and humanists). We consider that Humanists UK's point was therefore already covered by our drafting and have made no further change.

Dwellings and public order offences

A18.41 In its response, one stakeholder, [✂], noted that some of the Public Order Act 1986 offences do not necessarily apply inside a dwelling.[737]

Our response

A18.42 We set out our arguments for considering the dwelling point to be irrelevant to content judgements in our November 2023 Consultation. As stated, we did not consider the dwelling defence is likely to be relevant for online services, since content is routinely viewed on smart phones. Absent authority on this point, we have decided to keep our position the same.

Child sexual exploitation and abuse (CSEA): Offences relating to child sexual abuse material (CSAM)

Link between CSAM offence and IIA offence

A18.43 In its response, NSPCC (National Society for the Prevention of Cruelty to Children) argued that it would be valuable for the ICJG to acknowledge the connection between CSAM and the offence of sharing an intimate image without consent.[738]

Our response

A18.44 When it concerns intimate image abuse (or self-generated indecent imagery) of children, this will always be CSAM and the CSAM offences have a lower threshold for illegality than the intimate image abuse offence ('IIA offence'). For the purpose of service providers making an illegal content judgement about an intimate image of a child, service providers should therefore consider the CSAM offences. However, we acknowledge the link highlighted to us in NSPCC's response and have decided to add a steer for service providers to consider both the IIA offence and extreme pornography offences in cases where the subject of an image can be reasonably inferred to be over 18 and there are reasonable grounds to suspect that an image might amount to any of these two offences.

A18.45 For the purpose of carrying out a risk assessment, service providers should note that when a child is sharing self-generated indecent imagery, there may also be a risk of intimate

[737] [✂]
[738] NSPCC response to November 2023 Illegal Harms Consultation, p. 50.

image abuse offences manifesting on the service, in addition to the CSAM offences. We have added wording to this effect to the section titled "Risk assessment and illegal CSAM content" of the CSAM ICJG chapter.

AI-generated CSAM imagery

A18.46 NSPCC argued that Ofcom should work with service providers to ensure that AI-generated CSAM is being removed as illegal content.[739]

Our response

A18.47 Our guidance contains a clear steer that AI-generated content is within scope of the definition of both indecent images of a child (where it is lifelike enough to be classed as a 'pseudo-photograph') and prohibited images of a child (where it is not a pseudo-photograph). We also state that "Discussion of how to use generative AI for this purpose may also be illegal content if it amounts to encouraging or assisting the creation of such an image." We have therefore not made any changes to the existing text.

Inference of intent in CSAM offences

A18.48 In its response, the Canadian Centre for the Protection of Children (C3P) argued that it is important to distinguish between parameters for assessing content and parameters for assessing intent, as CSAM offences do not require inference of intent.[740]

Our response

A18.49 We agree that the requirement for CSAM offences is knowledge rather than intent. Our guidance does not ask providers to infer intent to identify CSAM.

Usage examples: prohibited images of a child (non-photographic)

A18.50 We have decided to make a minor change to one of the usage examples in this section of the chapter to add clarity around what type of content may amount to the prohibited images of a child (non-photographic) offences.

Usage examples: indecent images of a child

A18.51 Upon reviewing this draft ICJG chapter, we have decided to remove the following usage example: "A video recording, still from a video recording or photograph depicting a naked child where the genitals, anus or female breasts are visible, regardless of context or setting" from the guidance chapter. We have decided to do so because we are concerned that the description might be misinterpreted by providers to be not just an example, but a

[739] NSPCC response to November 2023 Consultation, p. 50.
[740] Canadian Centre for Child Protection (C3P) response to November 2023 Consultation, p. 31.

complete description of what an indecent image of a child comprises. However, a more specific example would risk being too specific to be helpful as an illustration.

Checking information with subject of an image

A18.52 We are aware that some offence areas are particularly at risk of malicious reporting from bad actors or perpetrators, including in the case of controlling and coercive behaviour.

A18.53 To lessen the risk of users using reporting and complaints maliciously, we give guidance that where a person other than the subject of the image itself states in a report or complaint that the potential victim is aged under 18 (or was aged under 18 at the time when the image was taken), service providers are encouraged to check this information with the subject of the image itself if they have the ability to do so.

A18.54 The same risk exists when it concerns malicious reports and complaints relating to grooming. We therefore give guidance that where a person other than the subject of the image itself states in a report or complaint that the potential victim is aged under 16 (or was aged under 16 at the time when the content was posted), service providers are encouraged to check this information with the subject of the image itself if they have the ability to do so.

Fraud and other financial offences

Hacking of accounts of public figures

A18.55 In its response, the Advertising Standards Authority stated it sometimes sees examples of legitimate accounts (including those belonging to for instance celebrities) appearing to be hacked to be used to post scam advertisements. It noted that a service provider is likely to have access to indicators regarding hacked accounts and suggested Ofcom should consider including this in the ICJG.[741]

Our response

A18.56 The Guidance applies to content posted by any account. If a legitimate account is hacked and posts content which amounts to fraud by false representation, then this is illegal content and must be removed.

Difficulty of identifying fraudulent content

A18.57 Stop Scams UK urged Ofcom to consider the unique difficulty in identifying fraudulent content. Fraudulent content is designed to appear as lawful content, and in many cases the advert or entry point for a scam journey will not be breaking the law at all. [742]

[741] Advertising Standards Authority response to November 2023 Consultation, p. 7.
[742] Stop Scams UK response to November 2023 Illegal Harms Consultation, p. 20.

Our response

A18.58 We acknowledge Stop Scams UK's point and have decided to add wording in the introductory section of the fraud chapter emphasising the difficulty of identifying fraudulent content. Our guidance warns providers to be aware that particular risks may be associated with accessing the link in question. Where service providers choose to follow links, it may be appropriate for them to use a URL checking service before doing so

Use of non-textual elements in content amounting to fraud by false representation

A18.59 In its response, UK Finance argued that the examples of indicators of fraud by false representation given in the ICJG are too narrow, and focus on written content rather than images, calls and livestreams.[743]

Our response

A18.60 As part of our decision to add information regarding how to risk assess for illegal content to each offence area, we specify that fraud by false representation content can take the form of any type of communication (including images and videos).

Offences relating to criminal property

A18.61 One stakeholder, [✂], recommended adding additional drafting to the section on criminal property, as follows: "Offences relating to criminal property include the sale of stolen goods and the sale of items which facilitate theft (in particular the advertisement/sale of electronic devices used to steal vehicles, which is being made an offence via clause 3 of the Criminal Justice Bill)." [744]

Our response

A18.62 We have decided to update the relevant section to state that "Offences related to criminal property include the sale of stolen goods and the sale of items which facilitate theft." However, the Criminal Justice Bill fell away due to the General Election in May 2024. We have therefore not added any detail about devices used to steal vehicles.

Technical corrections

A18.63 The Financial Conduct Authority (FCA) provided comments on, and some technical corrections to, the detailed descriptions of the financial services offences under the Financial Services and Markets Act 2000. In particular, the FCA highlighted its Perimeter Guidance Manual which provides guidance on the scope of regulation of financial services activity. [745]

---

[743] UK Finance response to November 2023 Illegal Harms Consultation, p. 16.
[744] [✂]
[745] Financial Conduct Authority (FCA) response to November 2023 Illegal Harms Consultation, pp. 9-10.

Our response

A18.64 We accept these corrections and have updated our guidance accordingly.

Other relevant laws

A18.65 The Competition and Markets Authority responded to our consultation flagging the importance of providers being aware that besides their responsibilities under the Act, they also have responsibilities under UK consumer protection laws.[746]

Our response

A18.66 We have added a reference to UK consumer protection laws, in particular the Consumer Protection from Unfair Trading Regulations 2008, to our guidance.

Definition of dishonesty

A18.67 We note that our account of false representation at consultation had not provided a definition of 'dishonesty."

A18.68 We have therefore decided to add a definition of 'dishonesty' at paragraph 6.40 of the ICJG.

Drugs and psychoactive substances

Prescription medication

A18.69 Greater Manchester Combined Authority argued that the ICJG should ensure the definition of drugs and psychoactive substances includes illegally sold prescription drugs. [747]

Our response

A18.70 A number of prescription medications are included in the list of controlled substances given in our chapter. We believe our drafting is therefore sufficient on this issue.

Provision of a list of controlled substances

A18.71 One individual agreed that an incomplete list of drug names is better than no list at all, and suggests that, given service providers may have their own lists, Ofcom collaborate with providers to create a common list. [748]

Our response

A18.72 We can confirm that we will seek to engage with service providers and others to keep our list up-to-date and to ensure that understanding of 'street names' and their chemical/legal

[746] CMA response to November 2023 Illegal Harms Consultation, p.1-2.
[747] Greater Manchester Combined Authority response to the 2023 Illegal Harms Consultation, p. 9.
[748] Fuller, A. response to November 2023 Illegal Harms Consultation, p. 26.

components is as widespread and commonly-held as possible. Since our November 2023 Consultation, we have added to our guidance some emojis which we are aware are commonly used to refer to drugs.

Monitoring of legal status of products commonly used to prepare drugs

A18.73 One stakeholder, [✂], noted that concern remains around the sale of some legal products used in the preparation of illicit substances online, most notably pill presses. It noted that envisaged changes under the Criminal Justice Bill may change the legal status of such products and the ICJG should be updated to reflect this if it is the case.[749]

Our response

A18.74 In the event that any changes to the law make it appropriate to review the ICJG, we will do so.

Illegal steroids

A18.75 In its response, the Center for Countering Digital Hate (CCDH) raised concerns about the prevalence of content advertising or promoting illegal steroids.[750]

Our response

A18.76 Steroids including nandrolone, testosterone and haolestin (fluoxymestrone) are controlled substances, as noted in the list provided in the ICJG. We recognise the significant risk posed by content promoting the use of such substances, and note that where content is 'offering to supply' them the ICJG clearly states that this should be removed as illegal content. We do not believe any change is required to the ICJG, but welcome the evidence the CCHD provided.

Firearms and other weapons

Inclusion of magazines in definition of 'component part'

A18.77 The Deactivated Weapons Association said that our description of 'component part' was inaccurate, as magazines are not considered a component part (under Section 57 (1D) of the Firearms Act 1968 and amended by Section 125 of the Policing and Crime Act 2017) and are therefore not controlled.[751]

[749] [✂]
[750] Centre for Countering Digital Hate (CCDH) response to November 2023 Illegal Harms Consultation, p. 2.
[751] Deactivated Weapons Association response to November 2023 Illegal Harms Consultation.

Our response

A18.78 As explained in our November 2023 Consultation, the priority offences in the Act include offences from every nation of the UK. To reduce the burden on providers, we have consolidated our guidance to the extent possible and where there are three very similar offences, have presented it as one. Magazines are included in the definition of "relevant component part" in article 37(1) of the Firearms (Northern Ireland) Order 2003. We have therefore included magazines in the definition of a component part. We have clarified in the chapter that the offences from Northern Ireland differ in this respect.

Shotgun ammunition

A18.79 One stakeholder, [✂] argued that it is not made sufficiently clear in the draft ICJG that it is an offence to buy or sell shotgun ammunition to anyone who is not a registered firearms dealer, as a result of the Firearms (Amendment) Act 1988.[752]

Our response

A18.80 At paragraph 8.44 (A6.40 in the draft ICJG) we set out is that is an offence to expose ammunition for a restricted firearm for sale if this is done by way of trade or business. We have added some further drafting to draw out that this includes shotgun ammunition.

Definition of an 'unauthorised person'

A18.81 A stakeholder, [✂] raised concerns that the language of an "unauthorised" person in the chapter on firearms is misleading, and risks confusion when used in relation to those legally able to sell/transfer etc. certain types of weapons such as shotguns and airguns. It argued that, while 'authorisation' is required to sell and transfer prohibited weapons, in the case of non-prohibited weapons what is required is that the person in question is a Registered Firearms Dealer.[753]

Our response

A18.82 We accept that using the terms 'authorised' and 'non-authorised' here could create confusion, and have therefore amended the ICJG to refer to 'Registered Firearms Dealers' and those who are 'legally permitted to sell, exchange or transfer the firearm in question.'

[752] [✂]
[753] [✂]

Sexual exploitation of adults

Addressing strategies to avoid detection

A18.83 In its response, Changing Lives referred to evidence showing that perpetrators make advertisements which invite prostitution (e.g. 'sex for rent' advertisements) vague in order to avoid detection and removal.[754]

Our response

A18.84 By definition, it is likely to be difficult in some cases to spot vague or purposefully misleading advertisements and so the ICJG will never be able to encompass all efforts to dissemble or hide true intentions. However, we welcome the evidence provided and have updated our guidance to note the common use of vague or suggestive language to hide true intentions.

Content which offers advice and support for independent sex workers

A18.85 Changing Lives noted that it would be useful for the ICJG to provide guidance on the difference between encouraging someone to engage in sex work and offering advice and support on how to engage in sex work safely. It noted that the majority of users posting advice and support in these instances would not gain from them becoming a sex worker.[755]

Our response

A18.86 We agree with this point and have decided to update our guidance to state that "(…) content which offers advice and support about how to engage in sex work safely is not illegal as it does not meet the threshold for intent to cause or incite another person to become a prostitute."

Clarity around scope of "services that are primarily or solely used for the purpose of selling sexual services"

A18.87 Changing Lives also asked for clarity about whether services such as Vivastreet fall within the scope of what the ICJG states are "services that are primarily or solely used for the purposes of selling sexual services." [756]

[754] Changing Lives response to the 2023 Illegal Harms Consultation, pp. 16-17.
[755] Changing Lives response to November 2023 Consultation, p. 16.
[756] Changing Lives response to November 2023 Consultation, p. 16.

Our response

A18.88 We have updated our guidance on the offence that relates to this response in a way that means this point is no longer relevant. Our guidance no longer uses the terminology described.

Concern regarding services dedicated to sex work

A18.89 In its response, Nordic Model Now! said services dedicated to sex work "normalise and legitimise prostitution and contribute to young people considering it a viable option."[757]

Our response

A18.90 We do acknowledge that some children and young people access adult services websites. However, the purpose of this guidance is to assist providers in identifying illegal content. We signpost to our consultation on age assurance for service providers publishing pornographic content, and to Ofcom's draft Children's Harms Guidance which provides guidance on what Ofcom considers to be pornographic content that is harmful to children.

Image-based sexual offences

Extreme pornography: "explicit and realistic" portrayal

A18.91 An individual argued that it would be valuable to add additional clarification to the explanation of "explicit and realistic" in the ICJG, noting that legislation specifically covers depictions, which includes scenes that are acted.[758] It is particularly important as in the past there has been confusion "that the law is only concerned with depictions of real events (particularly in relation to rape and sexual assault)."

Our response

A18.92 We agree that it is important to make this clear in the ICJG and have updated the paragraph in which we define "explicit and realistic" portrayals to include reference to acts that are simulated.

Extreme pornography: definition of pornography

A18.93 One individual noted that the definition of pornography set out in the draft guidance is incorrect. This definition states that "content can be assumed to have been produced either solely or principally for the purpose of sexual arousal" based on the nature of the image itself, not the intent of the uploading user or any viewer of it. The stakeholder argued that relevant legislation states that an image is 'pornographic' "if it is of such a nature that it must reasonably be assumed to have been produced solely or principally for

[757] Nordic Model Now! response to November 2023 Illegal Harms Consultation, p. 18.
[758] McGlynn, C. response to November 2023 Illegal Harms Consultation, p.17-18.

the purpose of sexual arousal." However, they argue that "for the purposes of a civil, regulatory regime, aiming to reduce the harms of extreme pornography, it would be appropriate to base systems, policies and practices on the basis of the nature of the image."[759]

Our response

A18.94 We thank the stakeholder for their response and agree that our approach is appropriate for the purpose of a civil, regulatory regime.

Extreme pornography: non-consensual sexual penetration

A18.95 Upon reviewing our draft guidance, we noted the need for a simpler and more understandable explanation of "non-consensual sexual penetration" with regard to the extreme pornography offence.

A18.96 We have therefore decided it would be helpful to add some clarity around what 'non-consensual sexual penetration' means and what service providers may want to consider when determining whether something posted is non-consensual. We also wanted to make sure the explanation of consent aligns with what we set out in the section on intimate image abuse. We have updated the ICJG accordingly.

Intimate image abuse: new priority offence

A18.97 In the draft guidance on the intimate image abuse offence, we set out that the Act would eventually replace the existing English/Welsh offence of intimate image abuse with a new, wider one.[760] We said that although at the time of writing the new offence had not been brought into force, the old one would be revoked and that the new one would become a priority offence if the Secretary of State decides to make regulations under section 222 of the Act, adding it to schedule 7. We therefore drafted the draft guidance assuming that the offence would be brought into force and would become a priority offence before issuing our final ICJG.

A18.98 The new intimate image abuse offence has now replaced the old one and is a priority offence, but because of the approach we took to this in the draft guidance, we have made no changes to the final guidance based on this.

Cyberflashing: harm of cyberflashing

A18.99 In our consultation, we noted that in many cases cyberflashing images are sent via direct messages and, in this context, we argued that the important thing is not so much that

---

[759] McGlynn, C. response to November 2023 Consultation, p. 17

[760] The new English/Welsh offence is multi-limbed. In our guidance we have focused on sub-sections 66B(1) and 66B(2). We have not dealt with sub-section (66B(3) as we believe that in practice, most if not all content which would be identifiable as amounting to this offence, would also amount to an offence under sub-section 66B(1).

providers remove the content (the recipient has already seen it and can delete it), but that victims have the opportunity to prevent further such messages being sent to them. We then went on to explain that as part of our proposed Codes measures, service providers at high risk of harassment are recommended to offer users the ability to block senders. We said we considered this tool would go some way towards enabling users to protect themselves from unwanted contact of all kinds.

A18.100 We received stakeholder feedback from an individual, who said that the wording around how the user can delete cyberflashing content misunderstands the significant harm of cyberflashing which "arises on being sent the unsolicited genital images in the first place". They argued that "the labour of constant deletion and blocking is part of the negative experience and harms."[761] In its response, Refuge also commented on the wording and argued that it minimised the harm of cyberflashing.[762]

Our response

A18.101 We acknowledge the harm caused by simply receiving unsolicited genital imagery. However, the ICJG cannot be used to recommend safety measures to providers; Codes measures would be needed to do that. Cyberflashing is also a non-priority offence under the Act, which means that the safety duty applies differently to it. However, providers have an obligation under the safety duty to take down illegal content of which they are aware. In our November 2023 Consultation, we were explaining our belief that this remedy would be unlikely to reduce the harm. While we understand that having to block users can be part of the negative experience or harm, we believe it is still important that, through our Codes measures, we recommend service providers put systems and processes in place which allows users to block other users from sending further content as it will help prevent further harm.

Cyberflashing: Usage examples

A18.102 In their response, an individual noted that one of the usage examples set out in the chapter which includes guidance on the cyberflashing offence refer to "causing alarm or distress" and argued that there is no requirement in the legislation to establish that the victim has been caused alarm or distress.[763] This stakeholder also argued that it might be more accurate to refer to genital images in one of the usage examples in the draft guidance which refers to an image or GIF. [764]

Our response

A18.103 We have updated the usage example to make it clearer that there is no requirement that the victim has in fact been caused harassment, alarm or distress.

---

[761] McGlynn, C. response to November 2023 Illegal Harms Consultation, p.16.
[762] Refuge response to November 2023 Illegal Harms Consultation, p.25
[763] McGlynn, C. response to November 2023 Illegal Harms Consultation, p. 19.
[764] McGlynn, C. response to November 2023 Illegal Harms Consultation, p. 19.

A18.104 We have updated the usage examples since the draft guidance, and we no longer refer to pornographic images or GIFS. However, we have added drafting in paragraph 16.17 of the ICJG to explain that the genitals being shown in the photograph or film being sent do not need to be the genitals of the user sending the content.

Unlawful immigration and human trafficking

Role of information provided by law enforcement

A18.105 SPRITE+ (University of Sheffield) expressed concern that heavy reliance on law enforcement information by service providers can lead to "collateral censorship," particularly when it concerns the inchoate immigration and human trafficking offences.[765]

Our response

A18.106 Please see Volume 3, Chapter 2, paragraphs 2.53 to 2.56 where we set our decisions in relation to the role of law enforcement in content judgements in more detail.

Coded language

A18.107 We are aware that posts amounting to the offence may use coded/vague language to avoid detection.

A18.108 As a consequence, we have decided to add drafting noting the likely use of such language at paragraph 12.5 of the ICJG.

Inchoate unlawful immigration offences

A18.109 A stakeholder, [✂], argued in its response that the 'inchoate or conspiracy' unlawful immigration offences can be committed via online content.[766]

Our response

A18.110 In the draft ICJG we set out how unlawful immigration offences may be encouraged or assisted online. In our final guidance on these offences, we have further set out how online content could amount to conspiracy to commit these offences.

Definition of exploitation

A18.111 When working through the final version of the ICJG, we noted that the definition of exploitation that we outlined in the draft guidance did not fully reflect the definition of exploitation set out in section 2 of the Human Trafficking and Exploitation (Criminal Justice and Support for Victims) Act (Northern Ireland) 2015 (c. 2(N.I.)) and section 3 of the Human

[765] SPRITE+ (University of Sheffield) response to November 2023 Consultation, p. 24.
[766] [✂]

Trafficking and Exploitation (Scotland) Act 2015 (asp 12). We have therefore updated the ICJG to ensure the description covers the broader Northern Irish and Scottish definitions.

Animal cruelty

Animals controlled by man on a temporary basis

A18.112 Born Free Foundation argued that Section 2(b) of the Animal Welfare Act 2006 ("[an animal … under the control of man whether on a permanent or temporary basis") should be fully included in paragraph A9.58 (15.12 in the final ICJG), as it referenced only permanent control.[767]

Our response

A18.113 We have updated our guidance accordingly.

UK jurisdiction

A18.114 We proposed that content is considered illegal if the cruelty occurs in the UK, is committed by a British person, or occurs in any country where it is an offence. Born Free Foundation argued this contradicts statements from Government that content should be removed if activity takes place outside the UK but is visible to UK users.[768]

Our response

A18.115 It is true to say that the Act applies to content where content amounts to a relevant offence within the UK illegal in the UK and visible to UK users. However, this interpretative rule in the Act applies only to what happens in relation to the content. It does not affect, for example, any offline circumstances required for the offence to be committed.

A18.116 In the case of the encouraging, assisting and conspiring offences, the animal cruelty offence being encouraged, assisted or conspired to etc. would need to be an offence which was somehow within the territorial jurisdiction of the UK courts. The rules the courts apply are very complicated and we do not consider that a service provider can proportionately be expected to be equipped to apply them. Our guidance broadly approximates some of the rules a UK court would apply in deciding whether an act was in territorial jurisdiction of the UK in a way which we think will be practicable for providers.

[767] Born Free Foundation response to August 2024 Further Consultation on Torture and Animal Cruelty, p. 5.
[768] Born Free Foundation response to August 2024 Further Consultation, p. 6.

A1.19 Online Safety Enforcement Guidance

Stakeholder responses by theme

Effectiveness of enforcement

A1.19.1 Glitch and South West Grid for Learning (SWGfL) said it was important for Ofcom to monitor and evaluate the effectiveness of its Guidance and enforcement actions over time and that the Enforcement Guidance did not set out how it would do this. [769]

Our response

A1.19.2 We aim to closely monitor the impact our enforcement action has on compliance with the requirements of the Act in general, particularly in the early years of the regime. Additionally, under section 178 of the Act the Secretary of State must review and publish a report on the operation of the regulatory framework, including the effectiveness of the enforcement powers, within five years of the last of the duties coming into force.

Flexibility

A1.19.3 One respondent, [✂], said it encouraged an Ofcom approach that recognises that services may be able to fulfil their compliance obligations through alternative methods and, if appropriate, demonstrate and evidence this through alternative methods. It highlighted the importance for services to have the flexibility to be able to comply with whatever methods work best for their service and its users.[770]

Our response

A1.19.4 One way service providers can comply with an online safety duty is to adopt the measures set out by Ofcom in the relevant Codes of Practice. However, as set out in section 49(5) of the Act, service providers can also choose to implement alternative measures to those in the Codes of Practice to achieve compliance. The Act therefore gives services the flexibility to choose the methods which work best for their service and users, provided they secure compliance with the provider's regulatory obligations.

Enforcing effective age assurance

A1.19.5 UKSIC said that when assessing whether children can access a service, Ofcom must verify that the age verification methods implemented by the service adhere to the standards outlined in the draft age assurance guidance[771] and that the enforcement process should

[769] Glitch response to November 2023 Illegal Harms Consultation, p.14; SWGfL response to November 2023 Illegal Harms Consultation, p.16.
[770] [✂]
[771] Ofcom, Consultation on draft guidance on age assurance and other Part 5 duties, 5 December 2023.

act as a mechanism to protect children from accessing harmful and age-inappropriate content.[772]

A1.19.6 Yoti emphasised the importance of Ofcom assessing the effectiveness of age assurance technologies relating to services putting in place mitigations to protect users from the most serious potential harms, or on mitigations that are relatively quick or simple to implement. The respondent also said there needs to be more granularity as to what the minimum standards for age assurance are. It also said it would be helpful for Ofcom to publish how it has determined that children have been able access a service where the service has deemed that impossible.[773]

Our response

A1.19.7 As set out in paragraph 3.3 of the Enforcement Guidance, securing a higher level of protection for children than for adults is something Ofcom is required to secure under the Act. The risk of harm to children is also a priority factor we will take into account when considering whether to take enforcement action, as set out in paragraph 3.9 of the Enforcement Guidance. Ofcom consulted on draft age assurance guidance for service providers in December 2023[774] and we plan to publish our statement in January 2025. In assessing the effectiveness of a service provider's age assurance systems as part of the enforcement process, we will consider the standards outlined in the age assurance guidance. If, as part of an investigation, we have determined that age assurance has not prevented children from accessing harmful content and this led to a breach of the requirements, we will publish our reasoning in any confirmation decision.

International enforcement

A1.19.8 One respondent, [✂], asked how Ofcom is planning to enforce against services based outside of the UK.[775]

Our response

A1.19.9 Service providers based outside of the UK are subject to the Act if the service they provide has a significant number of users in the UK, or if the UK is a target market. Many providers within scope of the Act are global businesses with a UK presence which provides an incentive for them to comply. In addition, as set out in section 7 of the Enforcement Guidance, we have tools that may enable us to take enforcement action against other members of the group, such as a subsidiary or parent company, even where a provider itself has no UK presence. We have a range of enforcement tools we may be able to deploy in cases of non-compliance, such as our power to apply for business disruption measures. Finally, we work closely with regulators in other jurisdictions which have similar online safety functions to Ofcom, including through the Global Online Safety Regulators Network, and have powers to co-operate with them, including disclosing information, to facilitate

[772] UKSIC response to November 2023 Illegal Harms Consultation, p.58.
[773] Yoti response to November 2023 Illegal Harms Consultation, pp.28-30.
[774] Consultation on draft guidance on age assurance and other Part 5 duties, 5 December 2023.
[775] [✂]

the exercise of their own online safety functions.[776] This may enable us to work collaboratively with an overseas regulator where it is better placed to take action against an overseas service provider.

Identifying service providers

A1.19.10 The Online Safety Act Network (OSAN) said that, in relation to small sites and message forums that sit behind URLs, the Information Commissioner's Office (ICO) has had experience of companies going into insolvency as soon as they are approached with regulatory measures. A different company is then set up with a forum named slightly differently until it is approached once again and carries on doing the same thing. The OSAN asked how Ofcom would determine who the service provider is in these cases and how it will keep track of them, as the Enforcement Guidance does not consider the issue.[777]

Our response

A1.19.11 The Enforcement Guidance sets out information on how Ofcom will normally approach enforcement under the Act, which has been informed by our experience and track record of enforcement in other sectors we regulate. We acknowledge that there may be challenges in identifying the appropriate service provider for a service in situations where companies fail to engage with us and act evasively to avoid any interactions. We intend to work with organisations and partners in law enforcement to help us identify the appropriate service provider in such cases.

Distribution of funds received from financial penalties

A1.19.12 Three respondents (including the Independent inquiry into Child Sexual Abuse Changemakers (IICSA), [✂], and UK Finance) recommended that any funds from financial penalties issued by Ofcom should be redistributed to support victims, including those of child sexual exploitation and abuse (CSEA)[778] and fraud.[779]

Our response

A1.19.13 Ofcom does not have authority over the distribution of funds received in response to the financial penalties we issue under the Act. We are required to pass these funds directly to the Consolidated Fund.[780]

Technology Notices

A1.19.14 The Internet Society said that the Act provides no guidance on processes and procedures for issuing a Technology Notice to an encrypted service. It said this meant these service

---

[776] Section 114 of the Act. We are able to exercise these powers in relation to the overseas regulators listed in The Online Safety (List of Overseas Regulators) Regulations 2024, which include the European Commission, the eSafety Commissioner in Australia and the online safety regulators in France, Germany and the Netherlands.
[777] OSAN response to November 2023 Illegal Harms Consultation, p.80.
[778] IICSA response to November 2023 Illegal Harms Consultation, p.2.
[779] [✂]; UK Finance response to November 2023 Illegal Harms Consultation, p.21.
[780] Section 400 of the Communications Act.

providers will not be able to foresee what they should be complying with on receipt of a notice and could not take any action to avoid receiving one. It asked that Ofcom provide clarification around the process and whether Ofcom would enforce a scan for criminal activity being facilitated by the encrypted service.[781]

Our response

A1.19.15 Information around when Ofcom will issue a Technology Notice and what the notice will require is set out in section 121 of the Act. We are consulting on separate guidance about how we propose to exercise this function, as required under section 127 of the Act.[782]

Criminal enforcement powers

A1.19.16 Three respondents commented on the use of criminal enforcement powers. Global Partners Digital was concerned about criminal liability for senior managers and urged Ofcom to trigger criminal liability as a proportionate last resort only.[783] The NSPCC requested more clarity in the Enforcement Guidance on how Ofcom will hold senior managers liable for compliance with CSEA requirements.[784] [✂].[785]

Our response

A1.19.17 As stated in paragraph 2.11 of the Enforcement Guidance, the Enforcement Guidance does not apply to the criminal enforcement of offences under the Act. We may publish further information on the use of our criminal enforcement powers in future.

[781] The Internet Society response to November 2023 Illegal Harms Consultation, 2023, p.9.
[782] See Ofcom's 16 December 2024 Consultation: Draft Guidance on the exercise of Ofcom's functions under Chapter 5 of Part 7 of the Act.
[783] Global Partners Digital response to November 2023 Illegal Harms Consultation, 2023, p.26.
[784] NSPCC response to November 2023 Illegal Harms Consultation, 2023, p.53.
[785] [✂]

A1.20 Guidance on content communicated 'publicly' and 'privately under the Online Safety Act

Stakeholder responses by theme

Location of individuals able to access the content

A1.20.1 The first statutory factor at section 232(2) of the Act sets out that Ofcom must consider the number of individuals in the UK who are able to access the content by means of the service.

A1.20.2 The Cyber Threats Research Centre at Swansea University suggested it would be useful if the guidance explicitly stated whether UK internet users are to be considered as unable to access content if a particular communication channel is geo-blocked in the UK but not elsewhere, notwithstanding the possibility that it might be accessed using a virtual private network (VPN).[786]

A1.20.3 The Institute for Strategic Dialogue suggested that it may not be possible for service providers to determine whether content is accessible to individuals in the UK on certain services (and to how many).[787]

Our response

A1.20.4 The Act allows for service providers to have different terms of service for UK users when compared to users elsewhere, but the online safety regime (including the first statutory factor) focuses on individuals in the UK. In the first instance, it is for providers to reach a view on how many individuals in the UK can access the content in question by means of their service.

A1.20.5 We recognise that there may be limitations to a provider's knowledge about the number of individuals in the UK that are in fact able to access content, given the information available to the provider and the need to make decisions about whether content is communicated 'publicly' at scale. We expect providers to make their assessment based on the information reasonably available to them, and on the inferences they may reasonably be expected to make from it. Where a provider is not able to precisely determine an individual's jurisdiction, we would expect the provider to adopt a sensible approach. A provider should

[786] Cyber Threats Research Centre, Swansea University response to November 2023 Illegal Harms Consultation, p.7.
[787] Institute for Strategic Dialogue response to November 2023 Illegal Harms Consultation, p.10.

be able to explain the methodology or assumptions it uses to determine the location of individuals in this case. We have not made any amendments to the guidance on this theme.

Status of metadata

A1.20.6 The NPSCC expressed concern that the guidance classes metadata as "private content" and urged us to reconsider this, given the important role that metadata plays to "detect and disrupt bad actors" on services.[788]

Our response

A1.20.7 For the avoidance of doubt, our view is not that *all* metadata should be considered as communicated 'privately' for the purposes of any proactive technology measures. The Act restricts our ability to recommend proactive technology measures to analyse both user-generated content communicated publicly and metadata *relating to such content*. Our guidance recognises this constraint, and we have therefore not modified it on this point. To clarify, metadata which relates to content communicated 'publicly' would not be considered as 'private' and could therefore be in scope of future proactive technology measures.

Replication in other markets

A1.20.8 Two industry stakeholders called for us to consider how the UK's approach might be replicated in other markets, especially those with less stringent legal protections concerning data protection, privacy, and freedom of expression. They argued that we should ensure that any regulations uphold the protections and expectations of UK users and safeguard rights globally.[789]

Our response

A1.20.9 We recognise that protections for users' rights vary in different jurisdictions. The Act makes clear, however, that the duties it places on Ofcom and on providers are in respect of the safety of UK users only. We are satisfied that the approach we have taken is in accordance with UK law and should further the interests of citizens and consumers in the UK. Therefore, we do not consider it appropriate to change our guidance in light of this stakeholder feedback.

[788] NSPCC response to November 2023 Illegal Harms Consultation, p.28.
[789] Apple response to November 2023 Illegal Harms Consultation, p.13; techUK response to November 2023 Illegal Harms Consultation, p.24.

Application of the distinction between content communicated 'publicly' and 'privately' to specific services

A1.20.10 In some stakeholder responses to the November 2023 Consultation, providers explained their understanding of the guidance – and the distinction between content communicated 'publicly' and 'privately' – as it would apply in the context of their own services.

A1.20.11 Wikimedia noted that it considers the distinction to be inapplicable to its services because their "public interest nature" means that any user-generated content is communicated 'publicly'.[790]

A1.20.12 [✂][791]

A1.20.13 Another service provider argued that the content on its service is generally communicated 'privately'. Its reasoning included that content on the service is not publicly searchable or discoverable.[792]

A1.20.14 In the first instance, it is for the provider of a service to which a proactive technology measure applies to determine whether content is communicated 'publicly' or 'privately' by means of their service. We have included some additional case studies within our guidance which will help service providers understand how we would likely approach a holistic assessment of the three statutory factors.

A1.20.15 [✂] we note that under section 232(3) of the Act, the following do not count as access restrictions:[793]

- a requirement to log in to or register with a service (or part of a service)
- a requirement to make a payment or take out a subscription to access a service (or part of a service) or to access particular content

A1.20.16 Therefore, the fact that a service uses a subscription-based model does not necessarily (by itself) mean that content on that service is communicated 'privately'.

A1.20.17 Even where there is an access restriction in place, we reiterate our position that this does not, by itself, mean that content is necessarily communicated 'privately'. We would expect the provider to consider the other two statutory factors where this is the case. Similarly, we reiterate our position (see paragraph 4.52), that where content is accessible by many people, it should be considered as communicated 'publicly' even where it may not be easy for individuals to discover.

---

[790] Wikimedia Foundation response to November 2023 Illegal Harms Consultation, p.24.
[791] [✂]
[792] [✂]
[793] [✂]

Application of the distinction between content communicated 'publicly' and 'privately' to specific services

A1.20.18 The Institute for Strategic Dialogue suggested that we should encourage providers to make clear to users which aspects of their service they consider are "more public or more private" and the consequences of this for users' privacy and the enforcement of their terms of service.[794]

A1.20.19 UK Interactive Entertainment (Ukie) gave the example of gaming companies, which it says already reminds players (for example, via codes of conduct or community guidelines), that they should not consider messages to other players to be "private" or "confidential".[795]

Our response

A1.20.20 We welcome service providers looking to provide clarity to users on which parts of their service are more public or private through reminders such as those referenced by Ukie in its response. However, we do not consider that providers should take such reminders or labels as determinative of whether content is communicated 'publicly' or not. The fact that the user labels content as 'private', for example, does not mean that the provider should automatically consider that content as communicated 'privately', without having considered each of the three statutory factors.

A1.20.21 Separately, we note that the Act requires service providers to include provisions in their terms of service giving information to users about any proactive technology used for the purpose of complying with illegal content duties (including when it is used). We have codified this requirement under measure ICU G1, as discussed in volume 2, chapter 10: 'Terms of service and publicly available statements'.

Accessibility

A1.20.22 The New Zealand Classification Office suggested that information on the types of measures that content communicated 'publicly' or 'privately' will be subject to, should be provided to users in a form that is accessible to the public, especially young people.[796]

Our response

A1.20.23 Volume 2 of this Statement sets out information on the types of measures that we recommend providers should take to mitigate the risks of illegal harm on their service. This includes measures that apply to both content communicated 'publicly' and 'privately', and to content which is only communicated 'publicly' (see volume 2, chapter 4: 'Automated content moderation'). We have also published an accessible summary of our recommendations, entitled 'Summary of our decisions'.

---

[794] Institute for Strategic Dialogue response to November 2023 Consultation, p.10.
[795] UK Interactive Entertainment (Ukie) response to November 2023 Illegal Harms Consultation, p.20.
[796] New Zealand Classification Office response to November 2023 Illegal Harms Consultation, p.7.

Wider use of proactive technology

A1.20.24 The Internet Watch Foundation (IWF) and BT Group suggested that some service providers might be confused as to whether proactive technology is permitted to be used to analyse content communicated 'privately'.[797] The IWF proposed that we should amend the guidance to clarify that providers can continue to use automated content moderation (ACM) on a voluntary basis.[798]

Our response

A1.20.25 We agree that service providers should have the option to use proactive technology in relation to content communicated privately by means of the service, including to detect illegal content, should they choose to do so. We have clarified this in volume 2, chapter 4: 'Automated content moderation'. However, we do not consider it appropriate to amend our guidance on the concept of content communicated 'publicly' or 'privately' to include this point.

[797] BT Group supplementary response to November 2023 Illegal Harms Consultation, p.1; Internet Watch Foundation (IWF) response to November 2023 Illegal Harms Consultation, p. 7.
[798] Internet Watch Foundation (IWF) response to November 2023 Illegal Harms Consultation, p. 7.