Additional Safety Measures

Annexes 13-15

Consultation

Published 30 June 2025
Closing date for responses: 20 October 2025

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

Contents

Annex

A13. Proactive Technology: Further evidence of relevants harms proposed as part of the measure

Introduction

A13.1 In Chapter 9: Proactive Technology, we set out our proposals for Proactive Technology Measures ICU C11, PCU C9, ICU C12, and PCU C10.

A13.2 The relevant harms proposed as part these measures are:

- illegal harms: image-based Child Sexual Abuse Material (CSAM), CSAM URLs, grooming, fraud and other financial services offences (fraud), encouraging or assisting suicide (suicide); and
- content harmful to children: Primary Priority Content (PPC), which includes pornographic, suicide, self-harm and eating disorder content.

A13.3 This annex sets out additional evidence regarding the benefits and effectiveness of proactive technology in detecting these relevant harms proposed as part of this measure.

Illegal harms

Child sexual exploitation and abuse (CSEA)

A13.4 As with the other harms discussed in the proactive technology chapter, human moderation alone cannot address the harm imposed by CSEA at scale. Addressing CSEA through human moderation also raises concerns around moderator wellbeing and safeguarding.[1] [2]

A13.5 We acknowledge that the majority of CSEA content may be communicated privately and that the proposed measures do not directly address this type of content. However, we consider that this measure can still significantly reduce harm for image-based CSAM, CSAM URLs, CSAM discussion and grooming.

A13.6 During an engagement workshop with people with lived experience of online harm, participants shared that they felt hash matching alone puts pressure on children experiencing harm to report an image of themselves before it could be actioned in the future. In our view, novel CSAM detection will help to reduce this pressure as images shared in public environments, including self-generated intimate images (SGII), are more likely to be automatically detected. One participant acknowledged that they are aware that

[1] See, for example: Spence, R, et al., 2023. The psychological impacts of content moderation on content moderators: A qualitative study. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *17*(4), Article 8. [accessed 19 February 2025]. Arsht, A., and Etcovitch, D., 2018. The Human Cost of Online Content Moderation. [accessed 19 February 2025].

[2] While this measure is likely to require an uplift in human review by providers, this can be minimised by automatic flagging and removal of 'repeat' images once it is verified the first time. We would encourage services to include moderator wellbeing as part of their considerations around moderation policies.

proactive technology exists, and suggested service providers should deploy it in instances where human moderation resource cannot meet demand.[3]

Unknown image-based CSAM

A13.7 Evidence indicates that the volume of CSAM is such that it would be unmanageable to moderate by human moderation alone. Thorn's automated classifier 'Safer' is used by providers to detect potential novel CSAM. They stated that over 1.5 million of the 3.8 million CSAM files detected in 2023 were predicted by the classifier as new or previously unreported.[4]

A13.8 Part of the effectiveness of these proposed measures is that they addresses two areas of CSEA that hash-matching technology[5] alone cannot capture: novel SGII and artificial intelligence (AI) generated images.

A13.9 We understand that effective proactive technology exists to detect or support the detection of unknown image-based CSAM, this includes automated content classifiers which prioritise content for review. We are aware of specific third-party providers that offer this (sometimes exclusively) for unknown CSAM. We are also aware of existing technology which provides a probability score that content may be AI-generated, which could be used for CSAM. Reality Defender stated their product can be used to detect CSAM including AI-generated face swaps and deepfakes.[6]

A13.10 In addition to this, newly detected content can then be hashed and added to databases, which can then be used to prevent the further spread of those images online for all those using hash-matching software.

Altered or disguised URLs

A13.11 Due to the measure's similarity with keyword detection, any evidence for the effectiveness of approximate (also called 'perceptual' or 'fuzzy') keyword detection and CSAM discussion is relevant for altered/disguised URLs, which is covered in the 'CSAM discussion' section (A13.15-18).

A13.12 For example, we are aware of existing approximate keyword detection tools that can be applied to custom lists of terms, which can include URLs. There is also existing technology that can detect URL redirections and shorteners.

Grooming

A13.13 We understand that grooming detection is complex, and detection methods have shown varying degree of effectiveness.[7]

A13.14 However, initiatives to detect this content exist. Providers may use metadata or combinations of signals from user profiles/content, in conjunction with AI, to help identify

---

[3] Ofcom/Lived Experience Workshop, 30 April 2025.
[4] Thorn, 2023, Thorn's 2023 Impact Report [accessed 18 March 2025.]
[5] See Ofcom's Illegal Content Code of Practice, ICU C9: Using hash matching to detect and remove CSAM.
[6] Meeting with Reality Defender, 19 December 2024.
[7] Gunawan, L. Ashianti, S. Candra and B. Soewito, 2016. Detecting online child grooming conversation. *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, Yogyakarta, Indonesia,pp. 1-6 [accessed 22 May 2025].

and classify grooming behaviour.[8] Natural language processing (NLP) is also a possible solution.[9] Providers may also be able to contribute to initiatives where signals of harm are shared between organisations which can then be used to investigate and/or enforce against similar activity on other services. This may assist in detecting grooming behaviour due to the cross-platform nature of the harm.

CSAM discussion

A13.15 Removing CSAM discussion involves removing CSAM-related terms and/or NLP to indicate potential CSAM-related conversations. CSAM keyword and/or discussion detection is a method that can be used to detect CSAM and is currently being used by industry to remove CSAM-related terms from services.

A13.16 Evidence suggests that keyword detection can be beneficial in reporting and removing content that may not have been identified in other ways (for example, through hashing technologies or other proactive technologies),[10] or may indicate where CSAM may be present.

A13.17 It may include text-based automated content classifiers to label content as indicative of certain CSEA-related behaviours, such as advertising or trading CSAM. There are also simpler solutions available that allow providers to integrate lists of coded terms into existing moderation processes to indicate the presence of CSAM.

A13.18 As suggested in 'altered or disguised URLs' (A1.11-A1.12), if a provider implements a keyword list as part of their solution, they can set this up in such a way that obfuscated terms are more likely to be detected as well.

Fraud and other financial services offences (Fraud)

A13.19 Fraud is a volume crime, therefore the need to deploy counter measures at scale is pertinent. We consider that proactive technology is likely to reduce user exposure to fraud at scale.

A13.20 In response to the November 2023 Consultation on Protecting People from Illegal Harms Online (November 2023 Consultation), several stakeholders highlighted alternative measures that are already in use or would be beneficial for services to implement. These included: URL detection;[11] image detection;[12] video detection;[13] machine learning

[8] Responses to our formal information request from [✂], Reddit, [✂], Yubo, and Pinterest. February 2025.
[9] P. Anderson, Z. Zuo, L. Yang and Y. Qu, 2019. "An Intelligent Online Grooming Detection System Using AI Technologies. *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1-6 [accessed 13 June 2025].
[10] Lee et al., 2020: Detecting child sexual abuse material: A comprehensive survey. Forensic Science International: Digital Investigation, Volume 34 [accessed 13 June 2025],
[11] Financial Conduct Authority (FCA) response to November 2023 Consultation, p.9; UK Finance response to November 2023 Consultation, p.16; Which? response to November 2023 Consultation, p.5-6.
[12] ABI response to November 2023 Consultation, p.2; UK Finance response to November 2023 Consultation, pp.2, 13. eBay response to the November 2023 Consultation, p.1
[13] ABI response to November 2023 Consultation, p.2.

classifiers;[14] AI;[15] red flag indicators;[16] and metadata analysis, including behavioural, data, technical signals and automated pattern analysis.[17]A number of providers have also publicly noted their use of machine-learning classifiers.[18]

A13.21 There is also evidence from services,[19] solution providers and academia suggesting that proactive technology can detect fraud effectively and accurately. Academic evidence suggests that advanced technologies such as machine learning classifiers, can be effective in detecting fraudulent content with high accuracy rates.[20] Similarly, solution providers have highlighted the improvements in the accuracy of fraud detection as a result of the use of behavioural analysis[21] in combination with machine learning.[22]

A13.22 While there are a number of off-the-shelf solutions to tackle this, some larger services will likely develop solutions in-house.

---

[14] FCA response to November 2023 Consultation, p.9; Google response to November 2023 Consultation, pp.43; Trustpilot response to November 2023 Consultation, p.24; UK Finance response to November 2023 Consultation, pp.2, 13; Which? response to November 2023 Consultation, p.10.

[15] Lloyds Banking Group response to November 2023 Consultation, p.5; Trustpilot response to November 2023 Consultation, p.24; UK Finance response to November 2023 Consultation, pp.2, 3; Which? response to November 2023 Consultation, p.10; eBay response to the November 2023 Consultation, p.1

[16] Cifas response to November 2023 Consultation, p.10.

[17] Reddit response to November 2023 Consultation, pp.10, 23; UK Finance response to November 2023 Consultation, pp.2, 13; Integrity Institute response to November 2023 Consultation, p.11

[18] Karen Hao, 2020. How Facebook uses machine learning to detect fake accounts | MIT Technology Review. [accessed 28 May 2025].; AirBnB, 2017, What we're doing to prevent fake listing scams [accessed 13 June 2025.]; Abhishek Chandak and Ritish Verma (LinkedIn), 2023, Augmenting our content moderation efforts through machine learning and dynamic content prioritization, [accessed 13 June 2025.]

[19] Google, 2024. Written Evidence to Parliament [accessed 13 June 2025.];
Ramos & Sbouai, 2025. Meta Launches AI Tool to Fight Scams on Facebook and Instagram | OCCRP [accessed 13 June 2025.];
Meta, 2024. Testing New Ways to Combat Scams and Help Restore Access to Compromised Accounts | Meta [accessed 13 June 2025.]; Amazon, 2024. How Amazon is using AI to spot fake reviews [accessed 13 June 2025.]; Twitch, 2023. Twitch State of Engineering 2023 [accessed 13 June 2025.]; Match Group, Commitment to Safety [accessed 13 June 2025.]; Tinder, 2024. Press Statement: Tinder partners with love island's Rob Rausch to help online daters watch for snakes and fakes [accessed 13 June 2025.]; Bumble, Bumble's A.I.-Powered Deception Detector Weeds Out Spam, Scam, and Fake Profiles [accessed 13 June 2025.]; Abhishek Chandak, 2023. LinkedIn: Augmenting our content moderation efforts through machine learning and dynamic content prioritization [accessed 13 June 2025.]; LinkedIn, Community Reports – Fake Accounts [accessed 13 June 2025.]; Airbnb, 2024. Airbnb's $200 Million AI Acquisition Is Redefining Your Next Vacation with New Tech Hires [accessed 13 June 2025.]; Seon, 2024 Behavioural Analysis in Fraud Detection. [accessed 13 June 2025.]; Infosys BPM, Fraud Analytics: How Behavioral Analytics Enhances Detection [accessed 13 June 2025.] ;

[20] Ramdas & Neenu, 2024. Leveraging Machine Learning for Fraudulent Social Media Profile Detection. Cybernetics and Information Technologies [accessed 13 June 2025.]; Atondo Siu & Hutchings, 2023. "Get a higher return on your savings!": Comparing adverts for cryptocurrency investment scams across platforms [accessed 13 June 2025.]; Gbivana, Okunola & Bright, 2025. Evaluating the Role of Social Media Data in Detecting Financial Fraud Patterns [accessed 13 June 2025.]; Bello et al., 2023. Analysing the Impact of Advanced Analytics on Fraud Detection: A Machine Learning Perspective, pp.103-126 [accessed 13 June 2025.];

[21] Seon, Behavioural Analysis in Fraud Detection. [accessed 13 June 2025.]

[22] Infosys BPM, Fraud Analytics: How Behavioral Analytics Enhances Detection [accessed 13 June 2025.]

Encouraging and assisting suicide (or attempted suicide)

A13.23 There is evidence of proactive technology being deployed to detect a wide range of suicide content (both PPC and illegal), and/or suicide proxies, both in-house[23] and by third party providers.[24]

A13.24 Academic evidence and research suggest that proactive technology such as machine learning classifiers can be effective at detecting suicide content.[25]

A13.25 Although it can be difficult to distinguish illegal suicide from primary priority suicide content, we recommend providers use proactive technology to detect both (assuming that they are at medium/high risk for both kinds of content). If proactive technology detects content that is likely to be either kind of suicide content, this should be moderated in accordance with existing content moderation measures (ICU 9 and ICU 10).

Protection of Children

Pornographic content

A13.26 We understand that it is current industry practice to deploy proactive technology for the detection of pornographic content. Human moderation has been known to lead to lasting psychological and emotional distress where moderators are exposed to disturbing content, raising concerns about the wellbeing of moderators.[26]

A13.27 In response to the May 2024 Consultation on Protecting Children from Harms Online (May 2024 Consultation), several stakeholders said that providers should proactively detect content harmful to children for moderation.[27] The National Society for the Prevention of Cruelty to Children (NSPCC) specifically said that without automated content moderation, services would be overly reliant on users reporting content harmful to children, which would be "entirely inefficient as a basis for identifying and protecting children from harmful content."[28]

A13.28 We understand that there is a range of proactive technology that may detect sensitive content depicting nudity and sexualised or inappropriate material. Although this may not

---

[23] Transparency Centre, 2025, Community Standards Enforcement | Transparency Center [accessed 13 June 2025.]; Transparency Centre, 2025, Community Guidelines Enforcement Report [accessed 13 June 2025.]; Snap Values, 2024, Snapchat Transparency Report | Snapchat Transparency [accessed 13 June 2025.];Meta, 2024, Preventing Suicide and Self-Harm Content Spreading Online | Meta [accessed 13 June 2025.]; Pinterest Policy, 2024, Transparency report | Pinterest Policy. [accessed 13 June 2025.].

[24] Moderation API, Self-harm model - Moderation API [accessed 13 June 2025.]

[25] Shaoxiong J., et al. 2018. Supervised Learning for Suicidal Ideation Detection in Online User Content. [accessed 17 April 2025]; Parsapoorm, M., Koudys, J.W., and Ruocco, A.C. Suicide risk detection using artificial intelligence: the promise of creating a benchmark dataset for research on the detection of suicide risk [accessed 17 April 2025]; Chatterjee, M., et al. Suicide ideation detection from online social media: A multi-modal feature based technique [accessed 17 April 2025].

[26] Bharucha, T. J., Lease, M., Riedl, M. J., Steiger, M., & Venkatagiri, S., 2021. The psychological well-being of content moderators: the emotional labor of commercial moderation and avenues for improving support. *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-14). [accessed 17 January 2025].

[27] VAWG Alliance response to May 2024 Consultation, p.12; The Commissioner Designated for Victims of Crime Northern Ireland response to May 2024 Consultation, p.5; The National Crime Agency response to May 2024 consultation, p.9; Centre to End All Sexual Exploitation (CEASE) response to May 2024 consultation, p.17 to 18; Vodafone response to May 2024 Consultation, p.2; and Nexus NI response to May 2024 consultation, p.15.

[28] NSPCC response to May 2024 Consultation, p.24

always align with the definition of pornography, we understand services may deploy technology depending on their terms of service, for example, where a service prohibits nudity.

A13.29 We know that a number of larger services currently deploy proactive technology to detect and action pornographic content or pornographic content proxy.

A13.30 In their transparency report, Meta Platforms Inc. (Meta) said that both Instagram and Facebook use automated tools to detect and action adult nudity and sexual activity. From April 2024 to June 2024, it actioned 11.9 million pieces of content on Instagram and 32.2 million on Facebook.[29] TikTok have stated that 31% of total removals from April 2024 to June 2024 were for "sensitive and mature themes", with 98.3% of this being proactive removals.[30]

A13.31 Snap Inc. has stated in its transparency report that of the 3,464,750 total enforcements and 1,822,215 total unique accounts enforced through the use of proactive technology, 1,291,158 enforcements and 596,352 accounts enforced were for "sexual content" with a median turnaround time (in minutes) from detection to final action of being less than one minute.

A13.32 Smaller services, on the other hand, may source external tools to detect pornographic content. Proactive technology tools we know are available include 'SightEngine',[31] 'Amazon Rekognition',[32] 'Microsoft Azure',[33] and 'PicPurity'.[34]

Suicide, self-harm and eating disorder content

A13.33 In response to the May 2024 Consultation, several stakeholders said that providers should proactively detect content harmful to children for moderation.[35]

A13.34 We know that several larger services currently deploy proactive technology to detect and action suicide, self-harm and eating disorder content or content harmful to children proxy.

A13.35 For example, Meta reported that it actioned 5.8 million pieces of content on Facebook and 6.1 million pieces of content on Instagram from July 2024 to September 2024.[36] In 2024, 19.1% of TikTok's total removals were for "mental and behavioural health", with 6% of this

[29] Transparency Centre, 2024, Community Standards Enforcement | Transparency Centre [accessed 13 June 2025.]
[30] Transparency Centre, 2024, Community Guidelines Enforcement Report [accessed 13 June 2025.]. However, it should be noted that this includes sexually suggestive content, nudity and body exposure, sexual activity and services, shocking and graphic content, and animal abuse.
[31] SightEngine. Detect nudity, porn, suggestive and explicit adult content in Images and Videos [accessed14 June 2025]
[32] Amazon Rekognition. Content-Moderation: say they can detect and label explicit images and videos, that their image moderation returns a hierarchical list of labels which indicate specific categories of adult content.
[33] Microsoft Azure, 2025, Adult content detection [accessed 13 June 2025.]. This provides adult content detection to detect adult material in images, their classification contains several different categories such as 'Adult images are explicitly sexual in nature and often show nudity and sexual acts'
[34] PicPurify, Porn detection in images - PicPurify [accessed 13 June 2025.]. This is an image moderation API which detects and filters images containing specific elements such as porn and nudity
[35] VAWG Alliance; The Commissioner Designated for Victims of Crime Northern Ireland; NCA; CEASE; Vodafone; and, Nexus NI responses to May 2024 Consultation.
[36] Transparency Centre, Community Standards Enforcement | Transparency Center [accessed 13 June 2025.]

in relation to its suicide and self-harm policy. Of the 19.1% of content removed, 91% was proactively removed.[37]

A13.36 Snap Inc. has stated in its transparency report that of the 3,464,750 total enforcements and 1,822,215 total unique accounts enforced through the use of proactive technology, 289 enforcements and 252 accounts enforced were for "self-harm and suicide", with a median turnaround time (in minutes) from detection to final action of less than 12 minutes.[38]

A13.37 Pinterest deactivated more Pins for violating its self-harm and harmful behaviour policy during this reporting period compared with H2 2023. It considered that this was due in part to introducing automated tools to action content for this policy in March 2024, and to increase matches of violative content identified and actioned by hybrid tools.[39]

A13.38 Yubo uses text moderation which covers various channels to detect risks such as self-harm or suicidal intention, while their visual moderation focuses on explicit signs of danger (e.g. blood, weapons). Their system relies heavily on pre-moderation to filter sensitive content.[40]

A13.39 Pinterest uses "keyword" detection, and 'maintains a list of sensitive terms and phrases' to 'block search results or prevent content from appearing in recommendations' where it may violate child safety, self-harm, suicide, drug abuse, and eating disorder-related policies.[41]

A13.40 Smaller services, on the other hand, may source external tools to detect suicide and self-harm, proactive technology tools we know are available. The availability of external tools to detect eating disorder harms is less well known.

A13.41 We understand through research and engagement with industry that external tools are available for the most extreme suicide and self-harm content. This may include those that detect blood, wounds, and graphic violence/gore more broadly. This may include text-based and image-based harm.

A13.42 Although we understand that proactive technology available for the detection of eating disorder harms may not be as widely available, we understand services may use behavioral identification and user profiling technologies to detect this type of content.

A13.43 Most consultations and statements will require an impact assessment, equality impact assessment and Welsh language assessment. Exceptions are explained in section 3 of the new Impact Assessment Guidance here. If assessments are not required, the consultation and statement should include a sentence or two explaining why.

A13.44 The impact assessment, equality impact assessment and Welsh language assessment can be included as sections in the main body of the document. They do not need to be presented as a separate annex.

[37] Transparency Centre, Community Guidelines Enforcement Report [accessed 13 June 2025.]
[38] Snapchat Values, 2024, Snapchat Transparency Report | Snapchat Transparency [accessed 13 June 2025.]
[39] Pinterest Policy, 2024, Transparency report | Pinterest Policy [accessed 13 June 2025.]
[40] Yubo response to our formal information request, February 2025.
[41] Pinterest response to our formal information request, February 2025.

A14. Perceptual Hash Matching for Intimate Image Abuse

A14.1 In this Annex, we set out the specific elements of our proposed measure for service providers to use hash matching technology to detect image-based intimate image abuse (IIA) content.

A14.2 The specific recommendations include:

- Using perceptual hash matching technology to detect image-based IIA content and conduct regular reviews of this technology to ensure precision and recall;
- Sourcing hashes from an appropriate third-party and/or internal database;
- Treating a match by the hash matching technology as reason to suspect the content may be intimate image abuse and moderate in accordance with our existing content and search moderation measures.

Conditions for the use of hash matching technology

A14.3 We are recommending that service providers use, where technically feasible, **perceptual hash matching technology** to analyse content to detect image-based IIA.

The content to be detected and analysed

A14.4 Service providers should analyse photographs, videos, or visual images that are generated, uploaded to, or shared on, the service after the hash matching technology is implemented. They should do this before or as soon as practicable after it can be encountered by UK users.

A14.5 Such content already posted on the service at the time the hash matching technology is implemented should be analysed within a reasonable time frame.

The configuration of the technology

A14.6 Service providers will have the **flexibility to set the threshold** to determine when there is a sufficient similarity between the hash in the database and the hash of the piece of content from the service. This flexibility is required to facilitate the fine-tuning of service providers' hash matching technology to produce the greatest accuracy in detecting and reducing the spread of intimate image abuse.

A14.7 However, we are recommending that service providers ensure the hash matching technology is configured to **strike an appropriate balance between precision and recall.** A perceptual hash matching system which seeks to find as much illegal content as possible (maximising recall) may result in an increased level of false positives (lowering precision).[42] Conversely, a system which seeks to minimise false positives (maximising precision) will detect less illegal content (lowering recall). In making a trade-off between precision and recall, the prevalence of illegal content is important. If illegal content makes up a low percentage of all content, a high proportion of detected content could be false positive results even with a technology which appears to have a low false positive rate because

[42] In the context of detecting matches for illegal content, a false positive is a case where the hash matching technology has incorrectly identified content as a match for illegal content (in this case, image-based intimate image abuse).

there are relatively few items of illegal content to be found, and many opportunities for the technology to wrongly identify other content as a match.

A14.8 The level of false positives (including any cases arising from content being incorrectly included in a hash database) determines the potential impacts on users' freedom of expression and privacy. These impacts are addressed in the rights assessment section of Chapter 11.

A14.9 We recommend service providers should consider the following when configuring the hash matching technology so that it strikes an appropriate balance between precision and recall:

- risk of harm from intimate image abuse, reflecting the service's latest illegal content risk assessment, and any information reasonably available to the provider about the prevalence of content that is intimate image abuse;
- proportion of content detected as a match by the hash matching technology that a false positive; and
- effectiveness of the systems and processes used to identify false positives.

A14.10 We are recommending that service providers review the balance between precision and recall at least every six months, to ensure that it remains appropriate.

A14.11 We are also recommending that providers keep a **written record of their approach** to configuring the technology and **review it at least every six months**. This promotes transparency and appropriate record-keeping of a service provider's use of hash matching technology and ensures service providers review and update this technology, if appropriate, based on the effectiveness and impact of the technology.

A14.12 To ensure the efficacy and accuracy of the hash matching technology, the content moderation systems and processes used by service providers to review detected content should ensure human moderators review and assess an appropriate proportion of content detected by the hash matching technology, taking into account the principles that:

- the degree of accuracy achieved by a service provider's automated systems and processes used as part of its content moderation function (as indicated by the periodic reviews and the outcomes of reviews of content carried out by human moderators);
- content more likely to be a false positive should be prioritised for review; and
- the importance of understanding the purpose and context of detected content when determining whether it is intimate image abuse content.

A14.13 Our recommendations for this technology review broadly align with our approach to the use of human moderators in the context of our existing measure (ICU C9) recommending hash matching for child sexual abuse material (CSAM).

Conditions for sourcing of image-based IIA hashes

A14.14 The success of perceptual hash matching will rely on the quality of the hash database used to source image-based IIA hashes.

A14.15 At this stage, we are not recommending that hashes are sourced from a person with expertise in the identification of image-based IIA content despite this being an element of measure ICU C9 (hash matching for CSAM).[43] We decided not to include this recommendation because the hash database ecosystem for intimate image abuse is not as

[43] Illegal content User-to-User Codes, ICU C9.7(a).

developed and operates differently compared to the ecosystem for CSAM. We understand that currently there is only one third-party hash database for intimate image abuse, and we do not want to create a barrier for service providers to implement this measure.

A14.16 Current third-party intimate image abuse database providers do not verify the content submitted for inclusion in their databases. The context-specific nature of intimate image abuse makes it challenging to verify because the images themselves are not illegal – it is the sharing of those images without consent or a reasonable belief in consent and/or intent to cause alarm, distress or humiliation that makes the content illegal.[44] Furthermore, to maintain the privacy of survivors and victims, and reduce potential barriers to use, current database providers do not have access to the images. Instead, a hash of the images is created by the survivor and victim on their device and shared with the database providers for inclusion in the database.

A14.17 It would therefore not be appropriate or proportionate for this measure to recommend the use of third-party databases with expertise in the identification of image-based IIA content.

A14.18 The proposed measure does **not prescribe the use of any specific third-party hash database**. Instead, we are setting out the following conditions for the selection of an appropriate set of hashes:

- **For a third-party database,** we consider it appropriate for providers to use either a verified or unverified set of hashes. An unverified set of hashes (or database) does not require hashes to be reviewed by an expert in the identification of image-based IIA content. Our understanding is that StopNCII.org (run by South West Grid for Learning) is the only hash database of intimate image abuse currently available to service providers and its database is unverified to protect the privacy of its users. However, other providers may enter the market over time. Service providers will also be expected to ensure that they obtain the most up-to-date version of the third-party database.
- **For an internal database,** providers can use an internal hash database of content identified as image-based IIA content. Internal databases may be pre-existing or developed in the process of implementing this measure. The internal database could be developed by including content detected via content or search moderation processes (including user reports or proactive detection) or by adding hashes from a third-party database.

A14.19 We consider that the choice and flexibility afforded to providers will ensure the long-term applicability of the measure. For example, given the proposed measure does not prescribe the use of a specific database, if more third-party databases of intimate image abuse are developed, service providers will be able to access them in accordance with the proposed measure. This is particularly important as we understand that currently the selection of third-party hash databases is limited and, therefore, this long-term flexibility is essential as the ecosystem grows.

A14.20 Service providers will need to ensure the hash database they use to identify intimate image abuse content on their services, either a third-party or internal hash database, is regularly updated with appropriate content.

A14.21 We are also recommending that appropriate policies are in place, and security measures taken in accordance with these policies, to **secure any hashes of image-based IIA held for the purposes of the measure** (including any copy of a hash database sourced from an

[44] Illegal Content Judgements Guidance (ICJG), para 10.36.

appropriate organisation). This is to protect against unauthorised access, interference, or exploitation. We consider mitigations for security risks may include but are not limited to:

- storing data securely within the service's system;
- restricting access to the intimate image abuse hash database to authorised persons only;
- maintaining records of all authorised persons;
- requiring multifactor authentication for access to an account capable of making changes to the intimate image abuse hash database;
- requiring that changes to the intimate image abuse hash database must be proposed and approved by more than one authorised person;
- avoiding the use of default or shared passwords and credentials for accounts providing access to the intimate image abuse hash database; and
- ensuring that passwords and credentials are managed, stored, and assigned securely, and are revoked when no longer needed.

Review of detected content

A14.22 For the purposes of this measure, service providers should consider a positive match by the hash matching technology as reason to suspect the content may be intimate image abuse.[45] This will trigger the existing illegal content moderation measures for both user-to-user and search service providers. This means relevant service providers will be expected to review the content to determine and either:

- make an illegal content judgement; or
- where the provider's terms of service or publicly available statement cover intimate image abuse consider whether the content is in breach of its terms of service.

A14.23 We set out information about how service providers should determine whether content is intimate image abuse in the Illegal Content Judgement Guidance.

A14.24 We are recommending that service providers' content moderation systems and processes should ensure that human moderators review and assess an appropriate proportion of detected content. We consider this to be a safeguard to identify content that is incorrectly detected as intimate image abuse and limit adverse impacts on users' rights.

A14.25 We are recommending that service providers should consider the following factors when deciding what proportion of content to review:

- the degree of accuracy achieved by the service provider's automated systems and processes used as part of its content moderation function (as indicated by the periodic reviews referred to above and the outcomes of reviews of content carried out by human moderators).
- The principle that content with a higher likelihood of being a false positive should be prioritised for review, with particular consideration regarding the use of an unverified hash database. An unverified hash database may include content that is not intimate

[45] To note, we recognise that it is not always possible for providers to identify whether content is image-based IIA, because context is needed to assess whether the images are being shared without consent.

image abuse and, consequently, could produce matches with online content that is not illegal.

- The importance of understanding the purpose and context of detected content when determining whether it is intimate image abuse.

A14.26 It is not always possible for providers to identify whether content is intimate image abuse, because context is needed to assess whether the images are being shared without consent. To minimise the resource burden of human moderation, we consider that once a service provider has reviewed a given hash to confirm that it is likely to be intimate image abuse, further hash matches of the same content do not need to be reviewed. In these cases, we expect that service providers may use other systems and processes for automated content and search moderation. This could include, for example:

- using cryptographic hash matching to identify if content detected as a match by perceptual hash matching technology was an exact match for intimate image abuse;
- using more than one perceptual hash matching algorithm (considerably reducing the likelihood that each algorithm results in a false positive for a particular content); and
- using machine-learning classifiers to identify items of detected content that are more or less likely to be illegal content.

A14.27 Appropriate moderation action will differ based on the type of service.

a) **For user-to-user services**, once a service provider has determined that content is intimate image abuse (and thus is illegal content), or is in breach of its terms of service,[46] they should swiftly take it down (where technically feasible).[47]

b) **For search services**, once a service provider has determined that content is image-based IIA (and thus is illegal content) or is subject to moderation action in its publicly available statement,[48] they should take appropriate moderation action.[49] This means they should ensure it no longer appears in search results or it is given a lower priority in the overall ranking of search results.[50]

A14.28 If a service provider judges that there is no reason to suspect the content is image-based IIA, where possible the service provider should pursue the following actions:

- If the content matched against a hash from a third-party database, notify the third-party database that the content is not image-based IIA.
- If the content matched against an internal database, update the database by removing the hashed content.

A14.29 Where possible, service providers should add newly identified image-based IIA content to their internal database and remove low-quality hashes.

---

[46] The provider may do this where it is satisfied that their terms of service prohibit the types of illegal content (such as intimate image abuse content) which they have reason to suspect exist. For more information see paragraphs 2.47-2.51 of Chapter 2 of Volume 2 of the Statement: Protecting people from illegal harms online (December 2024 Statement), p. 14.
[47] Illegal Content User-to-User Codes, ICU C2.3.
[48] The provider may do this where it is satisfied that the types of content included in the publicly available statement are broad enough to cover the type of illegal content (such as intimate image abuse content) that it suspects exists. For more information, see paragraphs 3.45-3.48, Chapter 3, Volume 2 of the Statement: Protecting people from illegal harms online (December 2024 Statement), pp. 102-103.
[49] Illegal Content Search Codes, ICS C1.3.
[50] Illegal Content Search Codes, ICS C1.4.

A15. Further detail on economic assumptions and analysis

A15.1 This annex provides further information related to the economic analysis used to support our provisional conclusions for some of the measures assessed in this Consultation. We outline assumptions we have used to develop quantified cost estimates across several of the measures.

General cost assumptions

A15.2 We have made some general assumptions on costs, which apply to our analysis of many of the measures. These general assumptions are usually combined with other assumptions that are specific to each measure to determine the estimated costs of the measure in the chapter in the main body of the consultation. Any additional assumptions that are used in the cost analysis are described in the costs section of the relevant chapters.

Labour Costs

A15.3 We have used data from the Annual Survey of Hours and Earnings ('ASHE'), to develop our estimates for the labour cost required to implement some code measures.

A15.4 All quantified estimates of costs are provided in 2024 prices, unless otherwise stated. This is the most recent data available. [51]

A15.5 Our December 2024 Illegal Harms Statement ('December 2024 Statement') and our April 2025 Statement on Protecting Children from Harms Online ('April 2025 Statement') include cost analysis for some similar measures and used 2023 ASHE data.

A15.6 To develop estimates for labour costs, we have used the ASHE 2024 gross median full-time earnings for the three occupations listed below. These occupations are likely to develop and/or manage the systems and processes that in-scope services will need to have to comply with the regime[52].

A15.7 The three professions we have determined to be most relevant for the measures, and the relevant Standard Occupational Classification ('SOC') 2020 references are as follows:

- Programmers and software development professionals' salary (2134) to estimate the cost of 'software engineer' time used when developing our cost estimates.
- Database administrators and web content technicians salary (3113)[53] to estimate the cost of 'content moderator' time when developing our cost estimates.
- Business, median and public service professionals' salary (24) to cover a range of professional occupations that are employed at various online services and might be

[51] Office for National Statistics ('ONS'), 2024. Earnings and hours worked, occupation by four-digit SOC: ASHE Table 14 - Office for National Statistics. Data is provisional at time of writing. [accessed 19 May 2025]
[52] ASHE documentation does not explicitly state that gross salaries include bonuses, but our understanding is that the gross pay includes bonuses, tips and other payments.
[53] This four-digit SOC 2020 code (unit group code 3133) includes occupations such as content, chat, web, and website moderators as well as other occupations such as database administrators and web content. technicians.

required to implement code measures. This could be legal employees, operations, product managers and so forth.

A15.8 For some service providers median UK wage rates may differ from actual salary rates. This may be especially the case for larger service providers based in the US, who may have higher salary levels. The salary costs of some types of staff, such as software engineers with certain specialisms, may vary and may be considerably higher in some cases. To take account of this, we have calculated a higher salary estimate, which is double the value of our lower estimate.

A15.9 Conversely, some service providers may outsource some relevant work to locations where average pay is lower than the UK, which may reduce costs. To the extent this is the case, our salary range may tend to overstate costs.

A15.10 We applied a 21% uplift to the gross wage costs to account for non-wage labour costs, such as employers' National Insurance contributions.[54]

A15.11 Table A15.1 shows the 'low' and 'high' labour cost estimates for different time periods, including the 21% uplift, for each of the three occupations. The figures are based on annual labour costs and we have calculated the monthly, weekly and daily estimates.[55]

Table A15.1: Low and High Range – Estimates of Labour costs

Occupation	Low	High
	Annual labour cost estimates	
Software engineer	£64,736	£129,472
Content moderator	£40,342	£80,685
Professional occupations	£52,042	£104,084
	Monthly labour cost estimates	
Software engineer	£5,395	£10,789
Content moderator	£3,362	£6,724
Professional occupations	£4,337	£8,673
	Weekly labour cost estimates	
Software engineer	£1,421	£2,842

[54] ONS, 2021. *Uplifts from wages and salaries to total employment costs: a note on data*. ONS recommends dividing the 'employer's social contributions (D.12)' by 'wages and salaries (D.11)' to arrive at the uplift. Both series are published as part of the annual UK National Accounts: Blue Book time series. They provide economy wide estimates of D.11 and D.12 annually. At time of writing, the most recent data available is for 2023. [accessed 19 May 2025]

[55] When producing cost estimates for the measures, we have used resourcing estimates based on different time periods (e.g. days/weeks/months) suitable for the particular measure. The annual wages are derived from the ONS, 2024 Earnings and hours worked, occupation by four-digit SOC: ASHE Table 14 - Office for National Statistics, Table 14.7a Gross annual pay for full-time employees, 2024 revised estimates. Monthly, weekly and daily wages are all derived from this annual figure. The monthly wages are derived from dividing the annual wages by the number of months in a year (12). The weekly wages are derived by dividing the annual figure by 45.54. The daily wages are derived from dividing the annual wages by the number of working days in a year. We assume on average there are 228 working days in a year. This assumes people work 5 days a week and that there are 8 bank holidays and on average people take an additional 25 days leave a year. [accessed 19 May 2025]

Content moderator	£886	£1,772
Professional occupations	£1,143	£2,285
	Daily labour cost estimates	
Software engineer	£284	£569
Content moderator	£177	£354
Professional occupations	£229	£457

Source: ONS (2024), Annual Survey of Hours and Earnings. Includes 21% uplift. Calculations are performed based on the underlying median gross salary (the 'low' estimate, before uplift is applied) and then uplifted by 21%.

A15.12 For the measures that require input from senior management, we have used salary estimates from additional occupations. These include senior managers and senior leaders with an estimates annual labour cost of £121,000 to £182,000.[56]

Non-engineering Costs for System Changes

A15.13 Where system or other software changes associated with a measure involve a software cost, we typically match the amount of engineering time with an equivalent amount of non-engineering time for work carried out by people in professional occupations. This is to account for labour time that a business might need to spend on a system change, for instance, legal or project management.

Maintenance Costs for System Changes

A15.14 Where system or other software changes associated with a measure involve an initial cost, we have assumed an ongoing annual maintenance cost of 25% of the initial cost. These ongoing costs reflect work likely required to ensure the system continues to operate as intended. We have applied this assumption in the absence of actual information about the ongoing maintenance costs. We applied this assumption in our costing work for several of our measures in our December 2024 Statement and April 2025 Statement. We did not receive any feedback from stakeholders on this assumption and have therefore kept it consistent with this work.

Further detail on age assurance cost analysis

A15.15 We analysed the costs of highly effective age assurance in the April 2025 Statement. We have used the same analysis here, using the same data.

A15.16 This sub-section provides further analysis of costs which has been used to support our conclusions on age assurance measures, as set out in section 18, Highly effective age assurance in the Illegal Content User-to-user Codes.

A15.17 We discuss:

a) Our general cost assumptions for age assurance.

[56] This is based on simple assumptions we have made of £100,000 salary for a senior manager and £150,000 for a senior leader, which are then uplifted by the 21% for non-wage labour costs.

b) Direct costs to service providers. We consider that all direct costs are likely to depend on how a service provider approaches its implementation of the measures, but in all cases we consider that the main costs are likely to relate to:

   i) preparing to implement age assurance; and

   ii) implementing and operating a third-party age assurance method; or

   iii) building and operating an in-house age assurance method.

c) Indirect costs to services due to our requirements to implement age assurance.

Our general cost assumptions for age assurance

A15.18 We adopt several general assumptions to estimate costs. The cost estimates are illustrative and may not capture the full range of possibilities in practice. Providers could face different costs depending on their circumstances (e.g. any existing age assurance capabilities), the age assurance method(s) they adopt and how they implement their age assurance process more broadly.

A15.19 We have assumed that users will have to confirm their age for each service separately. We recognise that where an online service provider manages multiple services it may be possible to share the age credential of a user across more than one service, which may reduce direct costs for the service and friction on users. Reusability of age checks and/or interoperability of age assurance methods may become more widely available in future, for example, where a user can complete an age check that is valid for many service providers. This could reduce costs and make implementing and operating age checks more cost effective for more services, and more convenient for users.

A15.20 We have assumed that service providers have no existing systems in place that can facilitate age assurance. Where services already have systems to gate access for users in some way or to obtain user information that may be relevant for age assurance (e.g., a payment system for subscription charges), the costs of implementing age assurance may be lower than our estimates suggest. This includes cases where a service may also be subject to regulatory requirements related to age assurance in other jurisdictions.

A15.21 We assume that age checks are one-off. As we set out in the Part 3 HEAA Guidance, we have not set specific expectations for service providers to repeat age checks. However, service providers should determine whether repeated age checks are needed to meet the robustness criterion based on the features of their service and age assurance process. For example, service providers may decide to conduct an age check each time an unregistered user visits a service. We note that services which do not offer accounts or where users do not choose to create accounts, could incur higher costs. Service may also face higher costs if users repeat the age check, e.g., children who repeat the age check after turning 18. However, we understand that some age assurance providers offer volume discounts to services[57] which could reduce the overall costs of these age checks for services which have to repeat age checks for users.

A15.22 We have assumed that services apply age assurance to all users. In practice, some services may be able to only age assure a subset of users. Depending on the specific context of a service, this may significantly reduce costs compared to the estimates we present. For instance, if a service conducted age checks for 50% of its users, and its costs were primarily

[57] DSIT, 2024. Online Safety Act impact assessment, paragraph 190 [accessed 12 June 2025].

driven by unit costs per age check, then we would estimate its ongoing costs related to conducting age checks to be up to 50% lower.

A15.23 We recognise that our cost estimates are dependent on the assumptions we have made. In practice costs could be higher or lower, depending on how service providers have decided to comply with their online safety duties and implement age assurance.

Preparatory costs relating to the introduction of age assurance

A15.24 All U2U services are likely to incur some one-off preparatory labour costs relating to the preparation of adopting age assurance. These may include staff familiarising themselves with the measures and guidance, familiarising themselves with ICO guidance, researching and assessing the suitability of different age assurance options for their service, considering how to implement age assurance in a way that is highly effective, meeting the relevant criteria and having regard to the other principles (such as accessibility).

A15.25 Where a service provider decides to use a third-party age assurance provider, the procurement process is likely to involve some time and effort related to governance and budget processes, evaluation of providers and senior management engagement. For larger businesses with relatively complex governance and procurement processes, a formal tendering process could tie up internal staff's time and take significantly longer.

A15.26 Overall, these preparatory costs are likely to depend on the size and type of service and are expected to be larger for large services because of different governance processes but also the number of employees likely to be involved.

Costs associated with third-party age assurance methods

A15.27 There may be upfront costs linked to the age assurance provider setting up a client account to prepare the age assurance method for use, or in some cases, this charge may be part of an ongoing maintenance support service.[58] We recognise that these upfront costs may be substantial for larger, more complex services, e.g. if the existing service infrastructure needs adjusting or there are other complexities with linking up the third-party technology with the services' systems or data. For instance, a report commissioned by DSIT found an example of a 'large gaming organisation' which received cost estimates from a third-party age assurance provider that it would incur '2-3 months of 4-5 developer's time' to implement the technology on its service.[59] However, our overall assessment is that this cost is likely to be small for most smaller services. For instance, we understand that some third-party methods are developed with ease of integration in mind, meaning that connecting to a services' existing systems should be relatively easy and cheap.

A15.28 The service provider may also need to train some of its staff who work closely with the age assurance process (e.g., software engineers maintaining the running of the age assurance software) when the process becomes operational. We expect such costs would be relatively small and could be larger in an alternative approach where age assurance technology is developed and implemented fully in-house.

A15.29 The main cost component relating to third-party age assurance methods is the per-check cost, including the cost to check the age of existing users and new users on an ongoing

[58] For example, based on Yoti's price list data from May 2022, setting up an organisational account is £750 per organisation. GC-13 Yoti Age Verification Pricing (digitalmarketplace.service.gov.uk) [accessed 12 June 2025].
[59] DSIT, 2024. Potential impact of the Online Safety Bill [accessed 12 June 2025].

basis. These costs are likely to vary depending on the age assurance process and provider, as underlying costs and pricing approaches vary. According to DSIT's impact assessment of the Online Safety Act, some age assurance providers offer volume discounts to services requiring a large number of checks and discounted fees for small clients and start-ups in some cases,[60] while subscription-based verification packages often include a fixed number of checks for users.[61] DSIT's evidence indicates that price per check ranges from less than 1p to £1, depending on the provider and method used.[62]

A15.30 To illustrate what these costs may mean for a service, we set out cost examples for hypothetical services with a different number of users in the table below. According to the Government's impact assessment on the Online Safety Act, most per-check costs provided were 10p or lower.[63] Our approach reflects the variety of methods and prices available in the market, as well as uncertainty about how the market may evolve in future.

A15.31 We use a low estimate of 5p per check and a high estimate of 30p. Considering the prevalence of volume discounts, we expect a smaller service is more likely than a larger service to incur per-check costs closer to the high estimate. However, this is partly mitigated by large services being likely to face more substantial preparatory costs or requiring a significantly higher volume of age checks.

Table A15.2: Illustrative cost estimates of age checks via third-party age assurance providers*

Existing UK user base	New users each year	Age assurance for existing users (**one-off**)	Age assurance for new users (annual ongoing cost)
1,000	100	£50 - £300	£5 - £30
10,000	1000	£500 - £3000	£50 - £300
100,000	10,000	£5,000 - £30,000	£1,000 - £3,000
350,000	35,000	£18,000 - £105,000	£2,000 - £11,000
700,000	35,000	£35,000 - £210,000	£2,000 - £11,000
1,000,000	50,000	£50,000 - £300,000	£3,000 - £15,000
7,000,000	70,000	£350,000 - £2,100,000	£4,000 - £21,000
20,000,000	200,000	£1,000,000 - £6,000,000	£10,000 - £60,000

Source: Ofcom analysis

**Note: For existing UK user base of 100,00 and more, cost estimates have been rounded up to the nearest thousand. These illustrative examples assume a faster rate of user base growth, in*

[60] DSIT, 2024. Online Safety Act impact assessment, paragraph 190 [accessed 12 June 2025].
[61] DSIT, 2024. Online Safety Act impact assessment, paragraph 186 [accessed 12 June 2025].
[62] It is possible that due to inflation that these examples are now out of date. Publicly available per check prices are greater than the bottom end of this range, and in these cases, it is not clear who these prices would apply to. DSIT, 2022. DSIT, 2024. Online Safety Act impact assessment, paragraph 190 [accessed 12 June 2025].
[63] DSIT, 2024. Online Safety Act impact assessment, paragraph 190 [accessed 12 June 2025].

proportionate terms, for the smallest services (10% growth rate) and a lower rate for the largest services (1% growth rate).[64]

A15.32 We assume that our code measures will mean that services will incur a one-off cost of checking the age of their entire existing user base.[65] To estimate costs illustrative costs, we multiply the number of users by the per-check cost (for example, 100,000 users x 5p = £5,000).

A15.33 We also estimate the annual ongoing cost of carrying out age checks for new users. We make illustrative assumptions about the volume of new users, assuming a higher growth rate for smaller services (10%) compared to larger services (1%).

A15.34 For simplicity, we assume that ongoing age checks on new users will continue, and that: (a) the cost per check remains unchanged over time; (b) all checks for a service cost the same; and (c) the nature of the service does not influence the per-check cost. Table A15.2 sets out a cost estimate for these ongoing checks.

A15.35 Services may incur other costs including for example software licensing costs, training costs and data storage costs. In most cases we assume these would be included in the ongoing age check costs.

A15.36 Various testing and evaluation activities are recommended under our highly effective age assurance criteria. Where services use third-party age assurance providers, we expect that those third parties would carry out the bulk of these activities, which may limit further costs incurred by services. However, service providers would still be expected to maintain due oversight and understanding of any third-party testing and evaluation, as it is the service providers in scope of our age assurance measures who are ultimately responsible for ensuring that their approach to age assurance is highly effective. This may therefore require some staff time on an ongoing basis.

A15.37 Due to the fast-developing age assurance industry and emerging new verification tools the future costs of third-party age methods are uncertain. We think there is a significant likelihood that costs of age assurance will fall over time, as well as the possibility of interoperability of different solutions to increase in the future.

Costs of developing an age assurance method in-house

A15.38 For illustrative purposes, we have also considered what an age estimation method could cost to develop and run.[66] We assume that the initial phase of work phase may take at least six months, which includes the design, development, testing and deployment of age assurance software. Development time and costs are likely to vary by the approach taken. The estimates we present below are intended to provide an illustrative example of the broad magnitude of costs associated with developing a single in-house age assurance method.

A15.39 The main costs are likely to be:

[64] We had not presented illustrative cost estimates corresponding to existing user base of 1,000 and 10,000 in the April 2025 Statement. We derive these additional illustrative cost estimates using the same assumptions and methodology as for the remaining cost estimates.

[65] We recognise that in practice this may take place over time as some users may not use the service frequently. We also acknowledge that the requirement to undergo age assurance may result in some user-drop off.

[66] In this section, we use estimates of labour costs based on 2023 ASHE data, as in the April 2025 Statement.

a) One-off labour costs relating to the upfront expense of developing, testing, and deploying the software. This would include ensuring that the age assurance process met the four criteria set out in our guidance: technical accuracy (evaluating methods against appropriate metrics), robustness (evaluating methods in real-world conditions), reliability (producing reproducible results), and fairness (testing and training the method on diverse datasets).

b) Ongoing staff costs of monitoring, supporting, and maintaining of the age assurance model. This would include meeting recommendations related to reliability, including monitoring key performance indicators and rectifying issues related to unexpected or unreliable predictions.

A15.40 Our high-level indicative analysis in the context of a large business (which we consider the more likely scenario[67]), suggests that the upfront costs of staff involved in the relating to development, testing and deployment of an in-house solution could be in the region of many hundreds of thousands and potentially up to £1 million.[68] The total staff costs, including other non-technical expertise, e.g. legal, may exceed this amount. In addition to these costs, a provider may incur substantial one-off costs relating to acquiring relevant datasets for developing its age assurance method and one-off software/hardware costs relating to additional computational resources to develop and train its age assurance method, which may include cloud infrastructure and data security.[69] A large service may be able to use existing infrastructure and resources for the purpose of a new age assurance process, there is still an opportunity cost to this because these resources are not available for other uses.

A15.41 There would also be ongoing staff costs relating to monitoring, evaluation and maintenance, and there could be additional ongoing data costs if the method requires significant improvements and/or changes in the future. We estimate that these ongoing staff costs could reach £1 million annually or potentially more, depending on a service's approach. Our estimates are based on the same salary assumptions for upfront and ongoing costs. In practice, it is possible that some ongoing activities could be conducted by more junior staff on lower salaries, such that ongoing costs could be lower than suggested here.[70]

A15.42 As with our examples on third-party methods, these cost estimates are only intended to be illustrative and depend on the different assumptions we have made. The analysis above relates to the development of a single method. Where a service develops multiple

[67] For example, Google has appeared in a registry of providers approved by the Age Check Certification Scheme (ACCS), the UK's program for age verification systems. https://www.biometricupdate.com/202312/google-receives-certificate-for-facial-age-estimation-in-the-uk#:~:text=Google%20has%20received%20a%20certificate,restricted%20content%20in%20the%20UK [accessed 12 June 2025].

[68] We assume that the upfront costs are based on staff input on a full-time equivalent (FTE) basis for around six months from c.16 software engineers, while for the ongoing labour costs we assume require c.14 FTEs annually. Costs may increase if the age assurance method involves a particularly high level of expertise, e.g., machine learning. This may be an overestimate given that we expect services could use more junior staff for some model monitoring, maintenance, and support functions.

[69] A service developing an age assurance method is likely to require a cloud security solution that runs all the time and scans information regularly. Securing the data and systems is needed from the development phase but the service will continue to incur this as the systems and data need to be secured on an ongoing basis.

[70] The ongoing labour costs we assume require 14 FTEs annually.

methods for use as part of its age assurance process, the total costs are likely to be significantly higher.

A15.43 Any services seeking to develop age assurance methods in-house are likely to be relatively large, due to the substantial upfront costs relating to software development and testing. This could be more cost effective if a service anticipates a high volume of age checks over time and lower ongoing engineering costs compared to the alternative of using a third-party age assurance provider. Large services may also already have the necessary employees to develop age assurance methods, including those with advanced skills who may be required.

A15.44 To the extent that smaller services have the relevant capabilities to pursue an in-house approach, it is possible that they may be able to do so more cheaply than suggested by our indicative cost estimates (e.g. due to having simpler organisational processes and lower overheads in relation to the relevant activities).

A15.45 The service may also incur some one-off staff training costs after age assurance is deployed to users, but these are likely to be relatively small in comparison to the one-off and ongoing costs relating to developing and deploying age assurance approach in-house and will depend primarily on the number of people that need to be trained and how much training is required.