Accreditation for terrorism/CSEA content detection technology

Annex 7: Consideration of additional stakeholder responses to our December 2024 Technology Notices Consultation.

Statement

Published: 8th May 2026

Contents

Annex

A7. Consideration of additional stakeholder responses to consultation

A7.1 In order to provide our advice on minimum standards of accuracy to the Secretary of State, we have carefully considered all responses to our December 2024 consultation.

A7.2 In Sections 2 and 3 of the Technology Notices Advice Statement, we provide a short summary of key stakeholder responses on our policy proposals and explain how we have taken these views into account in our final advice. In this Annex, we set out additional stakeholder responses we received and our view on the points raised.

A7.3 This should not be considered as an exhaustive summary – for this purpose, readers are advised to refer to our website to see all the non-confidential consultation responses in full.

Further feedback on the audit-based assessment (ABA)

A7.4 In addition to the stakeholder feedback on the ABA summarised in Section 2 of the Technology Notices Advice Statement, stakeholders also shared other suggestions:

a) One stakeholder suggested that terrorism and CSEA content should be treated as two separate types of risks and that Ofcom should separate its research and advice accordingly. They noted how each harm posed unique detection challenges and has different available solutions, with terrorism content being especially difficult to identify due to a lack of standardised definitions and hash databases, and the need for a contextual, nuanced review to prevent over-deleting content.[1] Google also noted, in response to our independent performance testing (IPT) proposals, the fact that different providers and jurisdictions have varying definitions for 'terrorism content' and 'CSEA content'.[2]

b) Others emphasised important considerations for our approach to assessing the Robustness Principle.[3] For example, one stakeholder requested an explicit assessment of reliability across different types of services.[4] The same stakeholder also suggested Ofcom clarify how lower-end and older devices will be accounted for in a scenario where a technology operates less effectively on devices with less processing power or memory, and noted the risk that these issues could be exacerbated when a technology is developed by third parties unable to test on a specific service on which it will be

[1] [✂].
[2] Google response to December 2024 consultation, p.9.
[3] [✂]; NCA response to December 2024 consultation, p.1.
[4] [✂].

deployed.[5] The NCA similarly noted the need for technology to be able to scale up to cope with large demand and to adapt to changes in offender methodology.[6]

c) X suggested we implement safeguards such as methods for continuous evaluation, including performance on X in particular (rather than hypothetical services during the audit/testing phase).[7]

d) Cyacomb suggested that Ofcom has used the term "technology" in a way that differs from common usage.[8] It suggested that our use of the term (and, by implication, the scope of accreditation) focused on specific products or tools which 'package up' technology, but that technology often has a more general meaning. It explained that the conceptual difference is important and that Ofcom should consider whether its use of the term risks distorting the intention of the Act. In particular, it suggested that - if section 121 of the Act is read with a 'common usage view' of what the word "technology" means - it may be possible to accredit hash matching in general (rather than requiring each product solution that uses hash matching technology to be separately accredited).

A7.5 Others highlighted specific elements of the assessment process:

a) Some raised specific security considerations. One stakeholder suggested the ABA should explicitly assess security and disclosure risks posed by third party integration or information leaks,[9] while another stakeholder urged Ofcom to have regard to commercial sensitivities and commercial fairness during the ABA.[10] Those responding as part of Open Rights Group's '48 hours to tell Ofcom: Practice safe text' campaign suggested assessing the new risks and harms introduced by the technology itself, for example those posed by undermining end-to-end encryption,[11] and alongside other stakeholders highlighted specific risks posed by image-based removal tools which may be subject to hash inversion attacks - resulting in the recreation of images from hashed data.[12]

b) Multiple stakeholders raised considerations relating to testing datasets. The NCA recommended that testing datasets should have explicit benign-illegal content ratios when testing the minimum standards of accuracy and, in the context of IPT proposals, that testing should be effective on synthetic, partially-synthetic, indeterminate and real images. They also noted that using the same datasets that included real CSAM across technologies could help build trust in the assessment, and raised questions about who would supply, assure and store such datasets legally.[13] Videntifier raised how hash

[5] [✂].
[6] NCA response to December 2024 consultation, p.1.
[7] X response to December 2024 consultation, p.5.
[8] Cyacomb response to the December 2024 consultation, pp.1-2.
[9] [✂].
[10] [✂].
[11] ORG's 48 hours to tell Ofcom: Practice safe text' response to December 2024 consultation, campaign summary, p.3; Open Rights Group response to December 2024 consultation, p.2, p.4.
[12] Internet Society response to December 2024 consultation, pp.3-4; ORG's 48 hours to tell Ofcom: Practice safe text' response to December 2024 consultation, campaign summary, pp.2-3; Open Rights Group response to December 2024 consultation, pp.2-4; Name Withheld 2 response to December 2024 consultation, p.2.
[13] NCA response to December 2024 consultation, p.2, p.3.

matching technology providers should be mandated to use specific named industry databases and to report results against those databases, or provided with a standardised test database, and suggested Ofcom accredit specific high-quality datasets to be used alongside accredited hash matching technologies.[14] Another stakeholder also noted, regarding our IPT proposals, that assessment data will need to be representative of the full spectrum of harm, and that the assessment should consider how the scope of the detection technology might affect the classification – for example, a classifier built to detect sexual extortion may label examples of CSAM solicitation as 'benign'.[15]

c) Some noted how software updates should be considered. The NCA suggested that the Maintainability Principle should incorporate established procedures for the non-emergency updating of software to avoid it becoming obsolete if, for example, online service providers update their core infrastructure quicker than the accredited technology can be upgraded.[16] Cyacomb requested guidance over what sort of changes to technologies allow them to remain accredited, and what would take them outside the original accreditation.[17]

d) The NCA stated that there needed to be flexibility with accreditation and the context of the company to which the technology will be applied, suggesting that end-to-end processes may need to be assessed in their entirety to ensure accuracy is assessed effectively (as multiple tooling could impact results).[18]

e) Within a broader call for applicants to have flexibility regarding which evaluation metrics they submit as highlighted in paragraph 2.22 in the Technology Notices Advice Statement, Google noted that the provision of performance metrics specifically tied to UK users may pose practical difficulties.[19]

A7.6 Some stakeholders asked for more information regarding specific Objectives. For example:

a) [✂] stated there was no clarity on how "consistent performance over time" is measured or ensured, noting that it was important to be clear to developers how information demonstrating consistency can effectively be provided as part of the audit-based assessment.[20] Another stakeholder similarly asked to what extent the same objective intersects with the objective relating to quality assurance plans within the Maintainability Principle, in particular in regards to machine learning or AI solutions where data drift and concept drift are expected phenomena that require a model to be regularly maintained and updated.[21]

b) The NCA suggested that terminology of the proposed Development in a Secure Environment Objective seemed vague, and that publishing more detailed agreed standards would provide clarity and ensure consistency in data security guardrails

[14] Videntifier response to December 2024 consultation, p.2.
[15] [✂].
[16] NCA response to December 2024 consultation, pp.1-2.
[17] Cyacomb response to December 2024 consultation, p.5
[18] NCA response to December 2024 consultation, p.3, p.4-5.
[19] Google response to December 2024 consultation, p.4
[20] [✂].
[21] [✂].

across technology developers. It noted that there may be government standards available.[22]

c) The ICO welcomed the inclusion of the same Objective and noted that relevant documentation could include appropriate evidence of how data protection by design has been complied with under UK GDPR, and suggested further engagement with Ofcom to ensure documentation fully aligns with the assessment objective.[23]

d) One stakeholder asked how the same Objective will be assessed for technology that has intentionally been open sourced, and whether interpretability within the Fairness Principle is primarily relevant in technology deployments where human oversight is lacking and decisions are automated.[24]

Our response

A7.7 It is not clear whether the stakeholder's point about terrorism and CSEA content requiring different standards refers to the ABA specifically, the broader accreditation process, or another part of it such as IPT. That said, we nonetheless agree that terrorism and CSEA content are different, and that this needs to be considered when assessing the accuracy of detection technologies. Our research in Section 2 of the Technology Notices Advice Statement recognises the differences between identifying terrorism and CSEA content – and in particular that terrorism content can require nuanced contextual judgments. This research informed the position in our advice, which sets out standards that are deliberately principles-based, and which can therefore apply to both terrorism and CSEA content detection technologies. Indeed, the separation of terrorism and CSEA content was an element of our proposed approach praised by other stakeholders, such as the Marie Collins Foundation.[25]

A7.8 We agree the points raised in relation to the Robustness Principle have merit, and in many cases, we have already covered these points in our recommended Objectives or have adjusted the wording to more explicitly account for different aspects of Robustness. For example, the wording of the fourth Objective for the Robustness Principle has been updated to explicitly reference reliable operation across the different services for which it was designed, and it already included references to a technology's ability to operate reliably across "varying system capacity demands". Similarly, we have also published a detailed set of indicative questions and illustrative evidence in Annex 9, to demonstrate the kind of information we would expect to have regard to when assessing applicants' evidence of meeting each Objective. Within this, question one under the Reliable Operation Across Relevant Services, Devices, and System Demands Objective references the need for technology to demonstrate reliable operation across all relevant devices and within the deployment settings for which it has been designed.

A7.9 It is not clear whether X's suggestion of continuous evaluation on X in particular applies to the way in which the ABA is conducted, or how frequently the accreditation window should occur. In terms of the ABA itself, we have sought to build such safeguards into this audit, for

[22] NCA response to December 2024 consultation, p.1.
[23] ICO response to December 2024 consultation, p.8.
[24] [✂].
[25] Marie Collins Foundation, response to December 2024 consultation, p.3.

example the inclusion of the Maintainability Principle to provide evidence that the technologies performance can be maintained over time and across deployment contexts. In relation to the frequency of the accreditation window, we discuss the requirement for re-accreditation at specific intervals from paragraph 3.57 of the Technology Notices Advice Statement. Regarding performance on specific platforms, we do not consider it appropriate at the stage of accreditation to engage regulated services in the evaluation process, as at the stage at which we are required by the Act to accredit technologies, we will not yet know on which particular service that technology may be used. When it comes to issuing a Technology Notice, as outlined in our Technology Notice Guidance, we will consider whether independent compatibility testing is appropriate to inform our view on whether the specific technology is suitable for the platform in question, and will engage the relevant platform at that stage.

A7.10 We have carefully considered Cyacomb's challenge about how we have approached the term "technology" in the context of our Technology Notice powers. Our view remains however that it is both necessary and appropriate for Ofcom to accredit individual technological 'products' for the purposes of section 121 of the Act rather than, for example, kinds of technology at a more general level.

A7.11 Had it been the intention of Parliament that we be able to accredit kinds of technology in the manner suggested by Cyacomb, we would expect this to have been made clearer in the Act. We note, for example, that Ofcom has the power to include proactive technology measures in some of its Codes of Practice which recommend the use of a "kind of" technology,[26] but that section 121 of the Act does not use this language. It is also unclear to us how the interpretation suggested by Cyacomb would work in the context of Technology Notices requiring the development or sourcing of technology. Section 231(11) also appears to support our approach as it recognises that accredited technology under section 121 of the Act is an example of content identification technology.

A7.12 We also consider there to be good reasons for Ofcom adopting the approach that it is, rather than accrediting technology at a more general level. The minimum standards of accuracy have been designed to ensure that the technology developer has taken steps to assure the accuracy of the technology, in advance of Ofcom considering requiring its use in a Technology Notice. There is a risk in our view that the purpose of the minimum standards of accuracy could be undermined if we were to adopt the view advocated by Cyacomb.

A7.13 Regarding the security considerations, we believe the security-focused Objectives within the Robustness Principle already address risks relating to third-party integration, information leaks or commercial sensitivities. For example, Annex 9 provides more detail on the type of evidence that we would expect to receive relating to the Development in a Secure Environment Objective, including in question 3 the cybersecurity and secure coding practices relating to external dependencies such as technology developed or maintained by a third-party provider. Similarly, the risk of information leaks is considered in question 5 under the same Objective, where it references multi-party approval processes for sensitive actions. Regarding any risks introduced by the technology itself, accreditation is not about assessing the performance and risks posed by a technology in a particular deployment scenario (for example, an E2EE environment). These are the kinds of risks that we would

[26] Paragraph 13 of Schedule 4 of the Act.

expect to consider when deciding if it is necessary and proportionate to issue a Technology Notice to a particular provider, rather than at the point of accreditation.

A7.14 In relation to the feedback regarding testing datasets, we note that our advice on minimum standards of accuracy does not require the testing of technology against specific datasets or datasets with a fixed ratio of illegal to benign content. We note in this regard that:

a) Had we recommended the use of IPT as part of the minimum standards of accuracy, then this would have required technologies in the same category to be tested by Ofcom or a third-party assessor independently against the same dataset. Annex 13 of our December 2024 consultation therefore considered some of the matters that we would have considered when sourcing a dataset (including the need for there to be an appropriate amount of illegal content in the dataset). As we noted therein, we would expect any technologies being used for the detection of terrorism and/or CSEA content to be mainly operating in environments where there is predominantly benign data. This means the dataset that is used in any IPT would ideally reflect that likely reality, meaning that Ofcom would have to consider whether the datasets would need to be imbalanced. Whilst we do not necessarily agree that there should be a fixed ratio of illegal to benign data, we recognise the importance of carefully considering the composition of the dataset and ensuring there is an appropriate balance between illegal to benign content in such a case; and

b) The ABA that we are recommending does include, as part of the Technical Performance Principle, Objectives relating to Performance Metrics and Dataset Quality. Whilst the Performance Metrics Objective requires that each applicant for accreditation demonstrate that their technology has been comprehensively evaluated (and that it provide the results of such testing to Ofcom) it does not require that the testing undertaken by each provider be identical (including against the same dataset(s)). Similarly, the Dataset Quality Objective does not require that applicants for accreditation provide evidence that they have used specific named datasets (or that their datasets contain specific illegal content or a set ratio of illegal/benign content). However, our view is that this is appropriate and proportionate for the purposes of the minimum standards of accuracy. We are mindful that, even if Ofcom or a third-party mandated the use of specific testing databases (or testing only against databases with specific illegal content or fixed illegal/benign ratios) as part of the ABA, there would be no straightforward way to know if tests against such datasets were conducted honestly and without any pre- or post-processing that artificially improves a technology's accuracy (this is reflected, for example, in our Key Finding 7). In the same way, accreditation is designed for a specific technological product (as mentioned in paragraphs A7.10-12) rather than to recognise a specific database for use alongside any accredited technology. Instead, the Dataset Quality Objective within the Technical Performance Principle is intended to reflect that we expect technology developers to provide evidence of the quality and suitability of their testing databases with respect to their use case when applying for accreditation.

A7.15 We agree with the need for accreditation to be flexible in relation to the different deployment contexts. That is why we have designed a principles-based approach to accreditation to ensure that technology developers have taken steps to assure the accuracy of technologies deployable in different contexts. As explained in our Guidance, we

also agree that an assessment which considers specific deployment contexts is also necessary, but view this as working best at the point of considering whether to issue a Technology Notice, rather than during accreditation – as is required by the Act.

A7.16 Responding to Google's point about the practical difficulties of supplying metrics specifically tied to UK users, our focus is understanding the performance of the technology. Therefore, evidence does not need to be limited to performance for only UK users, and we acknowledge that evidence submitted may feature performance metric results in respect of non-UK users.

A7.17 In terms of how technological updates are considered, we agree that these should be taken into account. The indicative questions and illustrative evidence in Annex 9 include some of the ways in which evidence of updates could be expected as part of accreditation against the Objectives within the Maintainability and Robustness principles. For example, the Consistent Performance Over Time Objective contains illustrative questions that reference testing procedures and records of performance across various updates to evidence consistency over time. How software updates may affect re-accreditation requirements will be considered during the operationalisation stage.

A7.18 Regarding the requests for more information relating to the Objectives, the indicative questions and illustrative evidence in Annex 9 mentioned above provide further detail on the type of evidence that we expect to receive related to each Objective and how we expect to assess each Objective. They include detail on the Consistent Performance Over Time Objective and the Effective Quality Assurance (QA) Plans and Periodic Monitoring Objective, with the latter featuring a question on retraining criteria relevant to data and concept drift. Additionally:

a) In relation to the NCA's suggestion to refer to a government standard to clarify expectations, we did not identify such a standard that was applicable to our specific use case. However, we expect the underlying questions sent to accreditation applicants to clarify the kinds of evidence that will be needed for each of the Objectives, including the Development in a Secure Environment Objective. We note that the illustrative questions reference, for example, ISO/IEC Standards 27001 and 27032. In our view, it is also more appropriate to clarify expectations through the questions rather than the wording of the Objective itself, so that the questions and evidence can remain flexible and be refined as technology, capabilities and international/government standards develop.

b) With respect to the ICO's suggestion of continued engagement with Ofcom about whether relevant documentation could include evidence of compliance with data protection legislation, we welcome the ICO's input in shaping our advice on the minimum standards of accuracy. We also note that while evidence of compliance with data protection may be closely related, the purpose of accreditation should remain focused on assessing technology against minimum standards of accuracy.

c) For open-source technologies, eligibility will depend on the presence of a clearly identified owner or responsible party for development and ongoing maintenance. The illustrative questions in Annex 9 also provide guidance regarding interpretability and the types of evidence we expect to require for the Objectives within the Fairness Principle.

Broader expectations of Technology Notices

Human rights

A7.19 In addition to the human rights concerns discussed in Section 2 of the Technology Notices Advice Statement relating to the scoring and thresholds of the ABA specifically, numerous stakeholders raised a broader point regarding the intersection of accreditation with existing human rights laws and data protection regimes.

A7.20 Multiple platforms and advocacy groups emphasised different human rights declarations, treaties and laws, including the Human Rights Act 1998 and the ECHR.[27] Numerous respondents suggested the accreditation process did not do enough to consider human rights,[28] and called on Ofcom to embed a human rights impact assessment within the accreditation process to safeguard such rights.[29] For example, [✂] and the Internet Society cited the REPHRAIN report as an example of a human rights-centred framework, developed as part of the UK Government's Safety Tech Challenge Fund to evaluate CSAM prevention and detection tools in end-to-end encrypted environments.[30]

Our response

A7.21 Ofcom recognises the importance of human rights, and the need for them to be adequately protected. For our assessment of the impacts on human rights of our advice on minimum standards of accuracy, see our human rights impact assessment in Annex 2. For more detail on how we will consider human rights in the exercise of our Technology Notice functions more broadly, please see the Technology Notices Guidance.

A7.22 In relation to the specific views above about the intersection of accreditation with existing human rights, it is our view that the accreditation process already considers human rights:

a) Firstly, the minimum standards of accuracy have been designed to ensure that technology developers have taken steps to assure the accuracy of accredited technology, including to act as a guardrail to protect users' fundamental rights. If Ofcom is considering issuing a Technology Notice, the accreditation process ensures that only technology that meets these standards of accuracy can be required. As a result, all of the Principles and Objectives assessed as part of the accreditation process are ultimately relevant to protecting human rights.

---

[27] Internet Society response to December 2024 consultation, p.5; [✂]; ORG's 48 hours to tell Ofcom: Practice safe text' response to December 2024 consultation, campaign summary, pp.1-3; Open Rights Group response to December 2024 consultation, p.1; X response to December 2024 consultation, pp.2-4.

[28] Big Brother Watch response to December 2024 consultation, pp.1-2; Internet Society response to December 2024 consultation, p.1,p.5; [✂]; Name Withheld 2 response to December 2024 consultation, pp.1-2; ORG's '48 hours to tell Ofcom: Practice safe text' response to December 2024 consultation, campaign summary, pp.1-2; Open Rights Group response to December 2024 consultation, p.1; X response to December 2024 consultation, pp.2-4.

[29] Big Brother Watch response to December 2024 consultation, pp.1-2; Internet Society response to December 2024 consultation, p.1; [✂]; Name Withheld 2 response to December 2024 consultation, p.1; Name Withheld 5 response to December 2024 consultation, p.1; ORG's '48 hours to tell Ofcom: Practice safe text' response to December 2024 consultation, campaign summary, pp.1-2; Open Rights Group response to December 2024 consultation, p.1; X response to December 2024 consultation, pp.3-4.

[30] Internet Society response to December 2024 consultation, p.1; [✂].

b) Secondly, our proposed approach is broadly aligned with the REPHRAIN framework referenced by stakeholders. While it was developed for a particular content type and deployment context, REPHRAIN's evaluation objectives such as "Effective Performance, Robustness, and Scalability", "Fairness/Non-bias" and "Maintainability" are a close match with our advised approach. One difference is REPHRAIN's inclusion of a formal human rights impact assessment, linked to a specific envisaged deployment context. As explained in our Guidance and at A7.15, we agree that such an assessment is necessary, but view this as working best at the point of considering whether to issue a Technology Notice, rather than during accreditation – as is required by the Act.

Data protection and privacy law

A7.23 Multiple stakeholders suggested that any conflict with data protection and privacy legislation such as UK General Data Protection Regulations (UK GDPR) should be considered in the assessment of technologies to be accredited.[31] For example, X argued that the audit stage should consider the interaction with other legislative frameworks more thoroughly to allow developers to address concerns upfront, rather than waiting until Ofcom is determining whether to issue a notice.[32] [✂], X and the Open Rights Group campaign also noted the practical challenges of applying technology to only UK users without affecting compliance with similar data protection and privacy legislation in other jurisdictions.[33] The ICO also stated that it expects technologies to be designed and deployed in full compliance with data protection law.[34]

Our response

A7.24 We acknowledge stakeholders' concerns about considering conflicts with other laws and regimes (including with laws in other jurisdictions) during accreditation, particularly in relation to privacy and data protection law.

A7.25 However, at the stage at which we are required by the Act to accredit technologies, we will not yet know all the details necessary to properly consider the extent to which other regimes may impact upon the exercise of Technology Notice functions, such as on which service (and content) that technology is being considered for use. Our view therefore remains that it is appropriate to consider the implications of requiring the use of a particular technology on other legal frameworks when we are considering issuing a Technology Notice to a particular provider (rather than at the stage of accreditation).

A7.26 We note that the Technology Notices Guidance is consistent with this and has been revised to make clear that we expect any technology required to be used in a Technology Notice is deployed in accordance with relevant data protection law. Our guidance also makes clear that, where a service provider anticipates a Notice impacting their ability to comply with other legal regimes, they can and should make representations to us about such concerns

[31] [✂]; Name Withheld 2, response to December 2024 consultation, p.1; ORG's '48 hours to tell Ofcom: Practice safe text' response to December 2024 consultation, campaign summary, p.2; X response to December 2024 consultation, p.3.
[32] X response to December 2024 consultation, p.3.
[33] [✂]; ORG's '48 hours to tell Ofcom: Practice safe text' response to December 2024 consultation, campaign summary, p.4; X response to December 2024 consultation, p.2-3.
[34] ICO response to December 2024 consultation, p.2

in response to a Warning Notice and we would consider this before deciding whether to issue a Technology Notice (and, if so, what requirements to include). Indeed, we are required before issuing a Technology Notice in a particular case to consider the risk of the use of the specified technology resulting in a breach of any statutory provision or rule of law concerning privacy (including, data protection law).

A7.27 We also note that the accreditation framework and minimum standards of accuracy that we have recommended do consider, and are intended to be consistent with, data protection and privacy where appropriate. For example, whilst the purpose of the minimum standards of accuracy is not to ensure compliance with data protection legislation (and the ICO is the relevant regulatory authority for enforcing data protection law rather than Ofcom), Ofcom would reserve the right to not consider a technology for accreditation where it is found by a Court or a competent authority such as the ICO to have been developed in breach of UK data protection requirements. Further, we have engaged with the ICO in preparing our advice to the Secretary of State on minimum standards of accuracy, and note that the ICO stated that there is a "synergy between the minimum standard of accuracy and data protection law",[35] whereby a "robust minimum standard of accuracy should support compliance with the data protection fairness and accuracy principles and mitigate adverse impacts on individuals".[36]

End-to-end encryption (E2EE)

A7.28 In the December 2024 consultation, Ofcom did not take a view on the extent to which there is technology available that could be used to identify or prevent users encountering terrorism and/or CSEA content in any particular deployment scenarios, for example end-to-end encrypted environments.

A7.29 Many stakeholders expressed strong views about the use of Ofcom's Technology Notice functions regarding end-to-end encrypted content, and the impact that this could have on individuals' human rights. Ofcom has considered stakeholder responses regarding the use of accredited technology on end-to-end encrypted content in its Technology Notices Guidance Statement.

A7.30 Other respondents had specific concerns regarding how our accreditation scheme might interact with end-to-end encryption:

a) Many stakeholders included in the Open Rights Group's '48 hours to tell Ofcom: Practice safe text' campaign stated that "Ofcom needs to consider how it scores and considers the […] risks and threats that could arise from accrediting any scanning technologies", for example the risk of undermining national security, UK data protection law or other human rights.[37]

b) Similarly, Chayn believed "that the accreditation scheme proposed does not provide sufficient safeguards to ensure that Technology Notices will not introduce significant new cybersecurity risks for users", adding that the Technology Notice regime proposed

[35] ICO response to December 2024 consultation, p.3
[36] ICO response to December 2024 consultation, p.3
[37] Name Withheld 2, response to December 2024 consultation, p.1; ORG's '48 hours to tell Ofcom: Practice safe text' response to December 2024 consultation, campaign summary, pp.1-3.

does not strike the right balance between safety and protecting fundamental rights, and may jeopardise access to privacy-preserving measures like E2EE.[38]

c) [✂] noted how it would never be appropriate for Ofcom to assess technology for accreditation that would undermine E2EE, adding that "technologies that would undermine E2EE should be discarded automatically".[39] Similarly, X stated that accrediting technology for E2EE services was not appropriate, as "proactive scanning that involves identifying specific pieces of content cannot be achieved without compromising end-to-end encryption".[40] A stakeholder also referenced a report from the European Data Protection Supervisor on the European Commission's Regulation Proposal on Child Sexual Abuse Material, in which it states that it is "technically impossible to reliably" detect content in an encrypted setting while maintaining privacy.[41]

Our response

A7.31 During the Parliamentary process, the UK government underscored that Technology Notices should be subject to robust safeguards to ensure that users' rights are adequately protected. The concept of accrediting technologies against a minimum standard of accuracy was included in the Act as one such safeguard, intended to reduce the risk that legal content is wrongly identified by providing a comprehensive and robust understanding of a technology's accuracy such that it could be considered for requirement through a Notice.

A7.32 Ofcom's powers under Section 121 of the Act are intended to be broad, flexible and technology neutral - the Act does not introduce any inherent limitations on the types of technologies which may be accredited. The minimum standards of accuracy have been drafted with this in mind - they are principles-based standards that consider the accuracy of a technology in the abstract rather than in a specific deployment context, such as in encrypted environments. We therefore do not consider it appropriate to rule out accreditation of any particular types of technology at this stage.

A7.33 In any case, Ofcom must act in accordance with human rights laws and consider a range of matters, including the extent to which the use of a specified technology might interfere with users' right to freedom of expression within the law, the risk of its use breaching privacy law, and whether the use of any less intrusive measures could achieve a similar outcome.

These and other proportionality considerations are taken into account at the point of considering whether to issue a Notice when we understand the deployment context, in accordance with the Act, rather than during accreditation. This is because a full consideration of matters relating to proportionality and users' rights requires clarity on the specific service to which a Notice is being issued, the type of content being addressed, and which accredited technology might be required. Further matters which Ofcom will consider prior to issuing a Notice, as well as procedures for information gathering, obtaining

[38] Chayn response to December 2024 consultation, pp.2-3, p.6.
[39] [✂].
[40] X response to December 2024 consultation, p.5.
[41] [✂] - the EDPS' 'Seminar on the CSAM proposal: "The Point of No Return?"' report, page 3.

representations from affected providers, and appeals are set out in our Technology Notice Guidance.

Regulatory alignment

A7.34 Multiple respondents suggested that Ofcom note or align its accreditation with other applicable regimes or frameworks:

a) For example, Google noted that Ofcom should endeavour to align its evaluation of technology for accreditation with principles and metrics that platforms are already required to track and optimise for other online safety regimes. It goes on to reference what providers of intermediary services are already required to report for automated content moderation systems under Articles 15.1(e) and 42.2(c) of the European Union's Digital Services Act (DSA).[42]

b) The NCA also suggested that Ofcom should align with other online safety regulators who adopt similar accreditations for the same technologies, noting services will want to avoid having to use different technology in different regions, and that global consistency with accredited technology could encourage adoption and engagement.[43]

c) Similarly, Ukie suggested that we adopt a similar approach to the Australian Online Safety Act and the DSA to ensure that accreditation requirements align with the actual risks presented by different service types rather than applying uniform standards across all digital platforms.[44]

d) One stakeholder also encouraged Ofcom to note the way in which the National Institute of Standards and Technology (NIST) has operationalized its Face Technology Evaluations scheme, as well as its rapid turnaround times.[45]

Our response

A7.35 As outlined above, Ofcom acknowledges the value of existing regulations and frameworks, many of which have informed our advice to the Secretary of State on minimum standards of accuracy. In May 2023, we commissioned an external consultancy, PUBLIC, to produce research into how different UK-focused and international accreditation processes from a range of sectors had been developed, evaluated, and operationalised. This research, found in Annex 5, included case studies into 11 accreditation approaches across 5 industries.

A7.36 It is also important to note that this accreditation scheme is unique, and we are not aware of any other regulators who have attempted to accredit terrorism and/or CSEA content detection technologies, or set minimum standards of accuracy.

A7.37 That said, we are open to the use of existing regulatory requirements as evidence by applicants during accreditation. For example, the content moderation metrics required by the DSA (noted by Google) could also be useful evidence to prove adherence to the Principles and Objectives within our recommended minimum standards of accuracy.

[42] Google response to December 2024 consultation, p.3.
[43] NCA response to December 2024 consultation, p.3.
[44] Ukie response to December 2024 consultation, p.7, pp.9-10.
[45] [✂].

A7.38 We also recognise Ukie's point about tailoring requirements to actual risk, which has informed our approach to our wider Codes of Practice. However, the recommended minimum standards of accuracy are intended to provide a comprehensive and robust understanding of the accuracy of a wide range of applicable technologies at the point of accreditation, where the deployment context is not known. It is not clear to us how we could (or why it would be appropriate to) tailor the minimum standards of accuracy to reflect the different types of risks presented by different types of service. Importantly, accreditation is not the final stage. Before issuing a Technology Notice to a particular provider, we will consider if the use of specific accredited technology is necessary and proportionate for the service in question. In this regard, we note that the Technology Notices Guidance recognises at paragraph 3.4 that – even if we consider it necessary and proportionate to issue a Notice to one service provider, and the nature of our concerns in relation to another service are very similar (for example, about the prevalence of CSEA imagery on a service), it does not follow that we would consider it to be necessary and proportionate for us to require the use of that same technology (or any accredited technology) on that other service.

Technical readiness of the market

A7.39 In the December 2024 consultation, Ofcom did not take a view on whether there were technologies available that could meet the proposed minimum standards of accuracy.

A7.40 Several stakeholders emphasised the limitations of the current market for third-party safety technologies.[46] They noted barriers in the following areas:

a) Some stakeholders noted that automated detection systems for both terrorism and CSEA content risk producing false positives, often due to limits in identifying and understanding context.[47] For example, one stakeholder noted how "technologies are fallible", and referenced "multiple cases of inaccurate classifications and false positives in external hash databases" such as the GIFCT database, arguing the process of adding content to databases should be transparent with effective oversight.[48]

b) A stakeholder noted the lack of training data available to developers, particularly for CSEA content, creates significant barriers to creating new tools or improving existing processes.[49]

c) Cyacomb suggested that newer third-party safety technology providers may face barriers to passing accreditation relating to scalability, if the audit prioritised proof that a technology has scaled over its potential to do so.[50]

[46] [✂]; [✂]; [✂]; Name Withheld 1 response to December 2024 consultation, p.1.
[47] [✂]; [✂]; Name Withheld 1 response to December 2024 consultation, p.1.
[48] [✂].
[49] [✂].
[50] Cyacomb response to December 2024 consultation, pp.3-4.

d) A stakeholder also highlighted what it viewed as technical limitations for available technology. Livestreamed content was presented as a challenge for such technology, due to the complexity of analysing images in the frame and audio in real time.[51]

A7.41 Several stakeholders maintained that in-house technologies are the most effective.[52] Arguments included that these tools were already tailored to the specific harms and context of platforms, and could be easily combined with other internal tools, datasets and expertise. One stakeholder also referenced a recent UK government paper in support of this argument, in which it claimed providers had indicated that the most effective detection technologies were developed internally and according to the needs of the individual platform.[53] The NCA noted that accreditation could be complex for in-house solutions given that these likely comprise multiple technologies combined with human moderation, though they did not expand on why they thought we may accredit in-house solutions.[54]

Our response

A7.42 We note the views of some stakeholders about the current market limitations and readiness, and about the barriers to obtaining training and testing data in some cases.

A7.43 Regarding the risk of false positives, we have taken this into account in our advice on minimum standards of accuracy by, for example, mandating the reporting of the false positive rate as part of the Performance Metrics Objective under the Technical Performance Principle. We have also recognised the importance of evidence regarding false positives (and the risks associated with false positives) in the Technology Notices Guidance. Specifically, paragraph 3.7c) of the Guidance makes clear that we would expect to have regard to evidence regarding the false positive rate of the technology under consideration, and from paragraph 3.16 onwards makes clear that we would consider whether independent compatibility testing is appropriate to inform our view on whether a Notice is necessary and proportionate in a particular case.

A7.44 Linked to the above, and in relation to the concern about some databases containing false positives, this is a matter that we would expect to consider when reaching a view on whether it is necessary and proportionate to require the use of a particular accredited technology (and, if so, what requirements to include in such a Notice). We note in this regard that, as discussed in Technology Notices Guidance Statement, we would be able to require the use of specific databases in a Technology Notice, where appropriate.

A7.45 We recognise the difficulties that technology developers may face in procuring suitable training data, particularly for CSEA content detection technology. Indeed, this is reflected in Key Finding 6 in Section 2 of the Technology Notices Advice Statement. We note that we have sought to take account of this in our advice on minimum standards of accuracy. For example, the illustrative questions that we have included in Annex 9 (and which are relevant to the Dataset Quality Objective) recognise that the applicant for accreditation may not have access to the underlying training and testing data.

---

[51] [✂].
[52] [✂]; [✂]; [✂].
[53] [✂] - the UK Government's 2024 'Technology and Trust and Safety' report, page 30.
[54] NCA response to December 2024 consultation, p.3.

A7.46 Regarding scalability, we agree that technologies that should have the potential to scale but that have not already done so should not be unduly penalised by the minimum standards of accuracy. As such, we have changed some of our Objectives to focus less on deployment and more on testing – for example, the Reproducible Performance Objective within the Technical Performance Principle now outlines how assessment can be based on testing, and where relevant, deployment. We encourage applicants to submit all the relevant evidence they have regarding their technology, even if it has not yet scaled, so that we or an appointed third-party can thoroughly understand its capabilities – including its potential to scale.

A7.47 In relation to the technical and practical limitations, we note that suitable technology may not currently exist. We do not however need to reach a view on this at this stage and do not therefore look to do so in this document.

A7.48 Regarding the use of in-house technologies, as outlined in the Guidance, Ofcom must take several steps before issuing a Technology Notice, including commissioning a skilled person's report. The relevant matters we will ask a skilled person to advise on may include an explanation of the service provider's existing systems and processes to identify relevant content, and how (and where) accredited technology could be implemented alongside this, or information on the prevalence of such content on the service. As such, the efficacy of services' existing technologies is expected to be considered as part of any decision on whether it is necessary and proportionate to issue a Technology Notice to a particular provider. If existing in-house technology is found to be sufficiently effective in detecting terrorism and/or CSEA content, it seems unlikely that a Technology Notice would be necessary and proportionate. That said, were an in-house technological solution to be submitted for accreditation, we do not consider this to be any more complex than for an external solution, given the flexibility within the audit-based assessment and the focus of accreditation on specific technologies as opposed to content moderation processes as a whole.

Market impacts

A7.49 In the December 2024 consultation, Ofcom set out its view of the likely impact of its proposals in relation to (amongst others) safety technology developers, and the providers of regulated services. Our view was that:

a) As outlined in Annex 6-8 of the consultation, many of the potential impacts to the market of safety technology developers – including costs and competition – arise at the point of accreditation (and reaccreditation) or at the point of issuing a Technology Notice. We noted that we had not considered in detail the costs of our proposed approaches (i.e., the ABA by itself, or the ABA plus IPT) as accreditation is optional, and we are not therefore compelling technology providers to undertake this process.

b) Our proposed approach to the minimum standards of accuracy should not have any direct impacts on the providers of Part 3 services, including on small and micro businesses. This is because accreditation of technology against the minimum standards of accuracy would not necessarily mean that any regulated service providers are

required to use that technology (nor even that we are recommending its use by those providers).

A7.50 In response to the consultation, some respondents expressed concern about the cost of accreditation on technology developers,[55] especially for SMEs.[56] Some of their concerns related to the risk that accreditation may be too complex, disproportionate or overly-burdensome.[57] To address this risk, some respondents noted the potential to leverage existing compliance data to reduce compliance costs for applicants for accreditation,[58] or to allow a single accreditation to apply to multiple services with similar risk profiles and identical functionalities to reduce unnecessary duplication and administrative burden.[59]

A7.51 However, several stakeholders also highlighted the potential opportunities accreditation could create for technology developers:

a) Some highlighted how Ofcom could use accreditation to support or encourage developers. Examples included offering "support, guidance, or even a sandbox environment" to help developers understand if technologies not yet mature enough for accreditation might be accreditable in the future,[60] or a pre-accreditation stage for emerging technologies to help developers secure investment.[61]

b) The NSPCC and IWF suggested making accreditation a publicly recognised and respected certification or industry standard to incentivise innovation and developers submitting technologies for accreditation.[62] The IWF suggested this could be similar to the safety-oriented BSI Kitemark certificate.[63]

A7.52 One stakeholder raised concerns about the potential impact of Ofcom's approach to minimum standards of accuracy and accreditation on the providers of regulated services. Specifically, they noted the risk that accreditation might create an inadvertent oligopoly and the resultant inflated costs and vendor lock-in.[64] That same respondent suggested that Ofcom should prioritise the accreditation of open-source technologies to build on initiatives, such as ROOST, that expand access to non-proprietary safety tools and ease the cost burden on mid-sized and smaller platforms.

---

[55] Cyacomb response to December 2024 consultation, p.14; Google response to December 2024 consultation, p.12; [✂]; NSPCC response to December 2024 consultation, pp.2-3; Ukie response to December 2024 consultation, pp.8-9.
[56] Cyacomb response to December 2024 consultation, p.14; [✂]; Ukie response to December 2024 consultation, pp.8-9; Videntifier response to December 2024 consultation, p.2.
[57] Cyacomb response to December 2024 consultation, p.14; Google response to December 2024 consultation, p.12; IWF response to December 2024 consultation, p.5, p.10; Ukie response to December 2024 consultation, p.2, p.3, pp.6-7, pp.8-10; Videntifier response to December 2024 consultation, p.2.
[58] Google response to December 2024 consultation, p.3; Ukie response to December 2024 consultation, p.8, p.9, p.10.
[59] Ukie response to December 2024 consultation, p.7, p.8, p.10.
[60] Cyacomb response to December 2024 consultation, pp.13-14.
[61] IWF response to December 2024 consultation, p.2, pp.4-5.
[62] IWF response to December 2024 consultation, p.4, p.5; NSPCC response to December 2024 consultation, p.1, p.2.
[63] IWF response to December 2024 consultation, p.5.
[64] [✂].

Our response

A7.53 For information on how we consider the potential impact of our advice on the safety technology market – including costs, competition, and changes to services – see our impact assessment in Annex 2.

A7.54 The costs concerns raised by respondents were relatively general and about accreditation as a whole. We have taken account of these, by for example not recommending IPT, in part because of the costs for Ofcom and technology developers.

A7.55 Where respondents did make more specific suggestions with a view to reducing the costs of compliance with the ABA, we have sought to address these. For example, metrics required for compliance with existing regulations or regimes may be useful evidence to submit for accreditation (and this could reduce the compliance burden for applicants). Regarding the suggestion that a single accreditation can apply to multiple services, it is unclear whether this refers to a single accredited technology, or is interpreting accreditation as applying to services. To clarify, as mentioned in paragraphs A7.10-12, the accreditation process is designed for specific technology 'products', and the circumstances under which this technology can be required by certain regulated services through a Technology Notice is outlined in the Technology Notice Guidance Statement.

A7.56 Ultimately however, we recognise that there will be costs for applicants from accreditation against the ABA. This is inevitable in our view and is a consequence of the requirement within the Act for technology to be accredited before it can be considered for use in a Technology Notice. We have however advised the Secretary of State to set minimum standards of accuracy that are no more onerous than is required, and are recommending a principles-based framework partly because we consider it provides more flexibility to applicants and should enable them to provide the evidence that they already have, without being unduly prescriptive. Finally, as previously mentioned in paragraph A7.49(a), accreditation is voluntary, and we are not compelling technology providers to undertake this process.

A7.57 In response to the potential opportunities for technology developers identified by stakeholders:

a) We will carefully consider what support for developers applying for accreditation could look like once the Secretary of State has published the final minimum standards of accuracy. In Annex 9, we have also published illustrative questions which should help applicants for accreditation understand the types and quality of evidence that we would expect to see in order to meet the minimum standards of accuracy.

b) It is important to note that accreditation is not designed to function like the BSI Kitemark certificate. Ofcom is required to publish an annual report on the exercise of its Technology Notice functions and technology which is being considered, or has been accredited, as meeting minimum standards of accuracy. However, as mentioned in the Technology Notices Advice Statement, just because a technology is accredited as meeting minimum standards of accuracy set out by the Secretary of State, it does not necessarily mean that Ofcom will require its use by a particular regulated service. Neither does accreditation signify that Ofcom or any other regulator has 'approved' a

technology or endorsed its use. The accreditation process is only intended to determine whether a technology could be considered for requirement through a Notice.

A7.58 We have considered the concern raised by one respondent about the risk that the approach to minimum standards of accuracy (and accreditation) may impact the providers of regulated services by creating an inadvertent oligopoly. We understand the respondent's concern to be that only a small number of technologies are accredited against the minimum standards and that this could have adverse impacts on any providers to whom a Technology Notice is ultimately issued.

A7.59 We note that the risk of only a relatively small number of technologies being accredited exists irrespective of the approach taken by Ofcom in its advice to the Secretary of State and in its more general approach to accreditation. This is because accreditation against the minimum standards of accuracy is, as noted above, entirely voluntary. We have however sought to mitigate this risk in a number of ways. For example, we are recommending a principles-based framework that is suitable for the wide range of technologies that might be in scope of our Technology Notice functions and are not recommending the inclusion of IPT; this is partly in recognition of concerns that the inclusion of IPT may increase costs and complexity for applicants, and thereby deter applications for accreditation. We have also sought to recommend minimum standards of accuracy that are proportionate, and not unduly high.

A7.60 We recognise concerns about how this might impact Part 3 providers to whom a Technology Notice is issued. However, we note that our Technology Notice Guidance makes clear that**,** we will where appropriate give flexibility in a Technology Notice between different technologies if more than one accredited technology is potentially suitable. Our guidance also makes clear that we expect, when reaching a view on whether a Notice is necessary and proportionate in a particular case, to consider the financial costs of compliance with a Notice alongside a range of other matters.

A7.61 Regarding any prioritisation plans for the accreditation of specific technologies, these have not yet been established, and will be considered during the operationalisation of the accreditation scheme. As mentioned in paragraph A7.18c), the eligibility of open-source technologies will depend on the presence of a clearly identified owner or responsible party for development and ongoing maintenance.

Third-party as accreditor

A7.62 In the December 2024 consultation, we explained that accreditation may be conducted by Ofcom or by a third-party appointed by us. We did not take a view on the involvement of third parties in the accreditation process, nor whether any suitable partners had been identified.

A7.63 Numerous stakeholders raised points for consideration should a third-party be appointed to conduct the accreditation process:

a) The NCA noted that the technologies submitted for accreditation will be "complex and innovative" and require a high level of expertise.[65] [✂] recommended that any third-party should have a panel that includes domain experts.[66] Ukie also stressed that any third-party should be well-equipped to assess the effectiveness of different types of technology.[67] It added that the selection process for any third-party should be transparent, and that stakeholders should be consulted on the criteria for selecting such a body to ensure industry confidence in the process.[68]

b) Multiple stakeholders suggested that any third-party should be independent or impartial.[69] The NCA added that appropriate "checks and balances" should be put in place to identify and deal with issues of impartiality, suitability and conflicts of interest in a transparent manner.[70] Google suggested that any third party should be required to provide evidence it is not developing, or associated with any developers of, illegal content detection and moderation technology.[71]

c) Google also highlighted concerns about the confidentiality and security risks of accreditation arising from platforms or technology developers providing information about the methods and systems they use to keep their platforms safe, noting that these would be "significantly enhanced" if Ofcom were to delegate the evaluation process to a third party. These include the risk of bad actors evading content moderation efforts, obtaining sensitive information and trade secrets and exposing user information. Google recommended that, in addition to its suggestion above, any third-party must provide evidence that they are able to safeguard the access to and confidentiality of this information, including using technical controls.[72]

d) The NCA also highlighted that, where more than one third-party were appointed, Ofcom must ensure that they are all consistently accrediting to the agreed standards.[73]

Our response

A7.64 We have not yet reached a view on whether Ofcom (or a third party approved by it) should accredit technology against the minimum standards of accuracy. However, we agree with the relevance of these factors if a third-party were to be appointed to run the accreditation process, and note that similar points were made by stakeholders as part of the multi-stakeholder workshop in December 2023 (on which, see Annex 4).

[65] NCA response to December 2024 consultation, p.4.
[66] [✂].
[67] Ukie response to December 2024 consultation, p.9.
[68] Ukie response to December 2024 consultation, pp.9-10.
[69] Google response to December 2024 consultation, p.13; [✂]; NCA response to December 2024 consultation, p.4; Ukie response to December 2024 consultation, p.9; X response to December 2024 consultation, p.5.
[70] NCA response to December 2024 consultation, p.4.
[71] Google response to December 2024 consultation, p.13.
[72] Google response to December 2024 consultation, p.13.
[73] NCA response to December 2024 consultation, p.4.