Summary

Google welcomes the opportunity to respond to Ofcom's consultation on Additional Safety Measures and to continue engaging with Ofcom on this important subject. We are supportive of Ofcom's goal of ensuring technology is used effectively to help keep users safe from bad actors, while preserving the core benefits of online environments, including the ability of children and adults to express themselves openly and access useful information while preserving their privacy.

As Ofcom will know from our prior engagement, Google is an industry-leader in the development and use of automated technologies to detect harm and ensure that services are safer by design. At Google, we have been working on this for years, and are continuing to invest in technology to help us tackle illegal and harmful content at scale.

We have set out in this response a number of practical improvements that could be made to the draft Codes to ensure Ofcom's policy objectives are met and that the proposed measures are workable and proportionate. We have raised concerns about areas where the Codes become too prescriptive or disproportionate, to mitigate the risk that measures included in the final Codes stifle flexibility and innovation in the use of automated content moderation, in a way that either inadvertently harms or fails to materially improve user safety.

We make the following general points regarding the draft Codes of Practice before providing specific suggestions to individual measures in the table below.

Approach to additional safety measures

As an overarching point, we note that services have only recently begun complying with new online safety duties. We therefore consider that it would be premature for Ofcom to make any final decisions on, and implement requirements to introduce, extensive additional safety measures before reviewing the impact and effectiveness of the existing measures in place.

In order to ensure a sound basis for policy-making, evidence on the efficacy of existing measures must be considered before Ofcom seeks to impose extensive additional measures. This will enable any additional measures to be prepared with the benefit of lessons learned from the first round of measures, to avoid any potential pitfalls. Furthermore, service providers are already expending significant resources to comply with the regime's current requirements and iterative but substantial changes will add significant cost to this existing compliance work.

“Potentially” illegal content

Measure ICU E2 requires in-scope service providers to ensure that content recommender systems are designed and operated so that “*content indicated potentially to be relevant content on the basis of relevant available information is excluded from users’ content feeds*”.

Targeting measures at content that is “potentially” relevant content appears to create a new tier or category of content, which is not defined or referenced in the the Online Safety Act (the “**OSA**” or the “**Act**”). This creates uncertainty and an associated additional burden of review, which is fundamentally inconsistent with the scope of the Act and risks undermining users’ rights to freedom of expression. This measure and the associated category of content should therefore be removed. There are a number of reasons why:

- *Introduction of a new category of content and type of judgement beyond that envisaged by the Act:*
    - In passing the OSA, Parliament made a deliberate legislative choice to confine safety duties on service providers to illegal content, content harmful to children, fraudulent advertisements and pornographic content. Each of these categories of content were defined in detail.
    - In doing so, Parliament decided that the OSA would **not** extend to content that was lawful but potentially harmful to adults. In giving evidence at the House of Lords Communications and Digital Committee, Dr Beatriz Kira, Assistant Professor in Law at University of Sussex, described the goal of Measure ICU E2: “to reduce the spread of content that is not necessarily illegal but could be harmful” (q. 13 transcript available here). This is clearly against Parliament’s intent.
    - Parliament also specified the approach to be taken where a system or process of a provider involves a judgement about whether content is content of a particular kind, and in particular about whether content is illegal content (s.192 OSA).
    - Contrary to the above, Measure ICU E2 appears to introduce a new category of “potentially” illegal content that was not intended by Parliament to be captured. This measure risks introducing an unduly burdensome requirement on service providers to identify “potentially” illegal content, that was not the intent of the Act, creating significant uncertainty for service providers on the standard to apply when seeking to identify such content.
- *Requirement to carry out “two tests” and proactively monitor*: As drafted, the measure appears to require service providers to first judge whether content is ‘potentially’ illegal, and then to assess whether it is in fact illegal. The proposal appears to amount to a general monitoring obligation which is both technically infeasible to implement, and

directly contradictory to Parliament's intent. There should be no requirement to proactively monitor to identify 'potentially' illegal content.

  - From a technological point of view, it would be necessary to automatically monitor, assess and evaluate all content on the platform to first determine whether it was "potentially" illegal (though at this stage no illegal content judgement, as required by the Act, would be made and therefore potentially illegal content could remain on the platform but be excluded from recommendations) and then subsequently an illegal content judgement is made and illegal content is removed from the platform. Most providers will not make a dual assessment but will make one judgement of illegality.

  - The introduction of an additional step and the requirement to design an ACM system to make two judgements is impractical and has limited benefit. It compels service providers to design systems to make two separate and delineated judgments, according to novel, vague and undefined standards (i.e. "potential illegality" and "reasonable grounds to infer illegality"). Both of these standards will require complex evaluations and the first will be especially error-prone because of its vagueness and the lack of guidance on its meaning. This consequently risks a significant adverse impact on user rights given the requirement to, in effect, remove or block content, without sufficient and precise knowledge of its illegality. Both of these standards will require complex evaluations and the first will be especially error-prone because of its vagueness and the lack of guidance on its meaning. This consequently risks a significant adverse impact on user rights given the requirement to, in effect, remove or block content, without sufficient and precise knowledge of its illegality.

- *Limited effectiveness beyond existing measures***:** Service providers are already required to have a content moderation function that swiftly takes down content (including from a recommender system) where they have reasonable grounds to infer that the content in question amounts to a criminal offence. In particular, the requirement to take such content down "swiftly" causes confusion as to the circumstances in which a provider would downrank content rather than remove it. We are unclear as to what additional safety benefit is derived from requiring service providers to set a lower threshold for the identification of "potentially illegal" content and to downrank this new category of content in users' feeds.

- *Interference with users' right to freedom of expression:* Clearly, content that is "potentially illegal content" may be legal. The measure will therefore necessarily result in legal content being made less likely to be encountered by users, impacting users' freedom of expression, beyond what the Act intended. We do not agree that exclusion of content from users' content feeds, rather than removal, acts as a sufficient safeguard for users' freedom of expression given content which is not recommended widely is far less likely to be encountered than it otherwise would have, had no measure been introduced.

Proactive Technology

Google is supportive of effective content regulation and is aligned with Ofcom's goal of ensuring regulation helps keep users safe from bad actors while protecting the core benefits of online environments, including the ability of users to express lawful speech openly, access useful information and connect with one another. At Google, we have been working on this challenge for years, ensuring the right policies to protect our products and users, and using both computer science tools and human reviewers to identify and stop a range of online abuse. A mix of human review and technology helps us identify illegal and harmful content and enforce our policies, and we continue to improve our practices and remain committed to transparency through regular updates to our Community Guidelines Enforcement Report. We have not waited for new regulation before acting to keep our users safe. We are constantly improving and introducing new policy changes to support online safety and continuing to invest in technology to help us tackle illegal and harmful content at scale.

However we are concerned that the draft Codes risk expanding the way in which services use proactive technology in an unfocused and disproportionate way. The draft Codes dramatically expand the categories of content which certain platforms are expected to proactively monitor for. Version 1 of the Codes required proactive monitoring for CSAM (both image-based and CSAM URLs) , and an expansion to cover grooming, fraud, encouraging or assisting suicide, terrorism, intimate image abuse and all kinds of primary priority content effectively obliges platforms to monitor all content for all harms (both illegal and legal). We have the following concerns:

- *Proportionality:* The use of proactive technology must be carefully balanced with the need to protect users' rights to privacy and freedom of expression. This was recognised during the passage of the Act, with clear constraints applied to its use, for example a recognition that proactive technology cannot be applied to private communications.
    - We are concerned that in expanding the use of proactive technology to a host of further types of content (especially the various categories of primary priority content, which are not well-defined in law or guidance, and therefore capable of being interpreted very broadly), Ofcom is applying excessively onerous measures that are not proportionate to the risks it has identified and not appropriately balanced with user rights.
    - The requirement to identify such a wide range of content risks making platforms subject to a general monitoring regime, which was not the intention of Parliament (as per Lord Parkinson who confirmed that the Act "does not require general monitoring of all content"- Session 1, 27 April 2023), and is not required by the Act.
    - Some of the proposed measures also fail to meaningfully differentiate between services that are high risk for a particular type of content that is harmful to

children and those that are medium risk for the same content thereby undermining an appropriately risk-based approach to burdensome and intrusive obligations.

- *Effectiveness*: Furthermore, there are practical difficulties in relation to the use of proactive technology to identify some types of content, for example grooming or eating disorder content. This can be infeasible for certain more subjective content categories, particularly if they are interpreted broadly. For example, as Ofcom itself recognises in its guidance on content harmful to children, there are practical difficulties in distinguishing between harmful content of this type (e.g. relating to self harm) as against helpful resources aimed at helping users searching for recovery content (para 1.28, Ofcom's guidance on content harmful to children). Requirements to proactively identify this type of content risks leading to over-moderation, preventing children from seeing content that may be supportive for vulnerable groups. In practice, this measure simply requires service providers to accept a greater administrative burden by identifying on at least an annual basis that proactive technology could not feasibly be implemented and / or does not meet all the proactive technology criteria, for no appreciable safety benefit to users.

- *Impact on Global Speech*: As Ofcom are aware, technology aimed at proactively identifying violative content has to be developed at a global level. It is simply not feasible or proportionate to build classifiers locally, given this would mean developing classifiers aimed at proactively detecting unlawful content as defined by hundreds of laws in hundreds of jurisdictions. Instead, effective trust and safety work in this area is best carried out by developing technology that captures core types of harms that impact users globally. By mandating the nature and scope of these technologies, Ofcom is unintentionally making platforms choose either to develop less effective, UK-specific technology that will make users *less safe*, or apply UK standards to users worldwide.

- *Technical Infeasibility:* The draft Codes provide that proactive technology must only be deployed where it is technically feasible to implement the proactive technology on the service. Through the definition of "relevant content", Ofcom also recognises that the Act does not permit proactive technology measures to be included in Codes in relation to "content communicated privately". We would welcome Ofcom's explicit recognition in its policy statement that it is technologically infeasible to use proactive technology on end-to-end encrypted (E2EE) services and that service providers will therefore not be expected to use proactive technology on E2EE content. We would also welcome a recognition that there should be a default presumption that E2EE content is content communicated privately. It is not possible to implement proactive scanning of E2EE without significantly weakening or breaking encryption. More broadly, Ofcom should recognise that E2EE has a valuable role to play in protecting users from certain harms and upholding users' right to privacy: it protects both individual users and collective public safety, for example by protecting users from unlawful surveillance, identity theft, fraud, and data breaches. We therefore consider that Ofcom's objective should be to

encourage rather than undermine the use of E2EE on regulated services. It is also fundamentally important that Ofcom protects E2EE in designing the Codes and related guidance, in order to fulfil its own duties to exercise its online safety functions in a way that protects users' rights to privacy.

Hash-Matching

Hash matching is a powerful tool, and one that Google has long invested in. It reduces the burden on reporting users, and it helps alleviate the need for a "whack-a-mole" approach to removals. It is an effective method for identifying duplicates and near-duplicates of content that has already been reviewed and actioned per individual products' content and product policies. This not only helps reporting users, but also provides an efficient mechanism for prioritizing resources and directing them towards previously unseen / unreported images and videos.

We also know that issues such as CSAM cannot be solved by any one company alone, and we are therefore committed to tackling it with others in our industry and partners who are dedicated to protecting children around the world. For example, Google has made hash-matching tools, such as CSAI Match, available for more difficult formats for hashing, like videos for use by industry partners and NGOs, as part of our Child Safety Toolkit. This tool is available and specifically used for CSAM.

However, hashing is not a silver bullet, and not all hashing is created equal. Not all content types are suitable for both first-party and third-party hashing solutions, and not all third-party hash databases are equal in quality and actionability. For this reason, the current draft measures that require hash-matching for terrorism content (for some U2U services) and intimate image abuse content (for some U2U and search services) are overly prescriptive.

Also, we have serious concerns that the use of perceptual hash-matching is presented as effectively the only technological compliance solution which will result in platforms falling within the safe harbour. We therefore recommend that Ofcom avoid proposals which could disincentivise solutions that go 'above and beyond' the specific proposals, penalise platforms that implement alternate, more effective solutions, rather than providing principle-based rules that are inherently future-proofed.

The requirement for services to use hash-matching in this regard could result in the Codes disincentivising platforms from innovating and updating to more advanced tools to combat terrorism and intimate image abuse ("**IIA**") content. The requirements to implement and/or assess proactive technology under ICU C11/C12 achieve the same aim as these measures, and it is therefore unclear what specific benefit there is to including them. We therefore recommend that these measures are either removed from the Codes or language is added to reflect that these measures are only illustrative examples of how the safety objective can be achieved.

Further, hash matching can be a highly effective tool for one type of content – but more complicated for another type of content. For example, hash matching for CSAM often requires

little else aside from a hash and the shared taxonomy (e.g., A1, A2, B1, B2). This is because — in most cases — CSAM is easier (than other types of content) to determine as illegal simply from reviewing the content itself, with no other added context.

IIA content – on its face – often looks visually similar to consensual adult pornography. The difference is context: whether that imagery was taken or shared with or without consent. This is most often **not** something that can be determined by reviewing the image or video itself. We note that this is the only type of content that search services are required to deploy hash-matching in respect of, but that in the context of search services, it is even more difficult to determine whether or not the imagery was shared with consent, given the extremely high bar for the "moderation" of content. It is therefore particularly disproportionate to require search services to deploy this type of technology in circumstances where Ofcom has not provided sufficient evidence that such technology is capable of being effective.

Similarly, with terrorist content, context matters. An image of a member of a proscribed group holding up a flag may be "illegal" or "violative" only because of the context surrounding that image. In a research article or a news article or in a post condemning terrorist acts, it would be a mistake to remove such content, even if it appears in a hash database.

As set out above in more detail, we would also welcome Ofcom's explicit recognition in its policy statement that it is technologically infeasible to use hash-matching on end-to-end encrypted (E2EE) services and that service providers will therefore not be expected to use hash-matching technology on E2EE content.

Livestreaming

We agree that livestreams by young users can pose risks. Live streaming involves creating content and interacting with your audience in real time, which may be better handled by older teens and adults. However, we also believe that livestreams can offer benefits to young users. For example, research indicates that benefits can include deeper social interaction and communication, stronger community building, and skill and knowledge development.

YouTube takes a proportionate approach that balances these risks and benefits, in ways that are appropriate for the platform. We have always maintained a high bar for access to our live features - for example, prior to being able to mobile live stream for the first time we require creators to complete phone verification, observe a 24-hour waiting period, and meet minimum subscriber thresholds. We also agree that placing restrictions on live-streams by the youngest users may be an appropriate mitigation of associated risks. YouTube has already taken a risk-based approach on the basis of available evidence about child development, only allowing livestreams from users aged 16 or above.

However, we consider that the measures proposed by Ofcom should afford more flexibility. Proposals to limit the features available to any under-18 users broadcasting one-to-many livestreams risk placing significant limitations on older children's ability to interact with, and

benefit from livestream features. Although we fully recognise Ofcom's concerns in relation to child users, rather than taking a blanket approach of blocking certain functionalities for livestreams by all teens, Ofcom should look to reflect the more proportionate approach they have taken for other measures where safeguards and effective content moderation processes, rather than functional limitations, are adopted and often on a risk-based approach. For example, ICU F1 asks for services to prompt relevant users before their network is expanded rather than limiting the ability to expand the network in the first place. Introducing relevant safeguards such as this to livestreaming would be a more proportionate approach than blanket restrictions. We believe there is scope for any new measures to better reflect the risk profile and nature of the service. The simple presence of a one-to-many livestreaming feature which under-18 users can access should not necessarily bring services into scope of new measures; rather only those services with a specific risk profile should be within scope. We therefore encourage Ofcom to restrict this measure to services that are at medium or high risk of grooming, as this appears to be the main driver for including this measure.

We would further reiterate that service providers are already adopting measures to take down illegal content and the efficacy of those actions in relation to livestreams must be considered before Ofcom takes further measures. At present, Ofcom's evidence base for the risk of harm does not account for the raft of cross-service controls that will have been adopted this year.

Amendments to Illegal Content Judgement Guidance (ICJG)

Where the contents of a private group chat or forum are E2EE, Ofcom's proposed amendments to the ICJG recommend that service providers assess "connected content" (including the name, bio, or image of the group chat) and infer illegality on the basis of this connected content unless there is evidence to the contrary. We would welcome Ofcom's recognition that, where connected content is encrypted, it is not technically feasible to detect this content and service providers cannot be expected to make an illegal content judgement.

We note that there are significant consequences where service providers identify CSAM, including mandatory reporting and account enforcement. Given these consequences, service providers should be expected to have confidence that they have reasonable grounds to infer whether CSAM is present in order to balance the impact on user rights and efficiency for public authorities. A service's judgement based on an assessment of connected content may not meet this high bar of confidence and risks unfair impacts on users as well as "over-reporting" non-actionable instances to law enforcement, creating inefficiencies.

The proposed amendment is also disproportionate as it's unclear what data sources services should have to consider before they have confidence to make an inference that the content is CSAM. In the absence of clear guidance, service providers may be required to actively consider connected content, third party hashes, and external reports to make an inference of illegality which would represent a significant burden for services.

Your response

Question	Your response
Question 1: Do you have further evidence regarding the harms and risks to users from livestreamed illegal content or content harmful to children, or harms and risks to children from broadcasting livestreams?	Confidential? – Y / N
Question 2: Do you have further evidence regarding the benefits to users or children from livestreaming?	Confidential? – N We are concerned that Ofcom has not considered the economic benefit to users, and children that are older teens in particular, of livestreaming when making their decision to restrict children's access to certain livestream features such as monetisation or user engagement. Paragraph 6.59 of the consultation acknowledges that these monetisation and engagement impacts on 16- to 17-year-olds may be particularly severe. However, there is no quantitative assessment of the value of the streaming economy to those users and to the UK economy more generally. Assessing the potential economic impact of the proposal (not just the service provider's cost of implementation) is essential to considering the proportionality of the measure. By way of example, YouTube and Oxford Economics found that the creator economy contributed £2.2bn to the UK economy and supported over 45,000 jobs in 2024. While livestreaming by older teens will clearly only represent a proportion of this economic contribution it nevertheless demonstrates the significant economic risk that restrictions to the monetisation of and engagement with livestreams may represent. In line with broader government objectives to support growth, Ofcom should consider the economic benefit of livestreaming to both users themselves and the economy more broadly. Research indicates that benefits to children of livestreaming can include deeper social interaction and communication, stronger community building and skill and knowledge development. For example, real-time engagement with a creator can help create a

	genuine sense of excitement in users and foster a valuable online community.
Question 3: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Relevant proposals: ICU D17 (The provider should have a mechanism to enable users to report that a livestream contains content that depicts the risk of imminent physical harm.) and ICU C16 (The provider should, as part of its content moderation function (see ICU C1), ensure that human moderators are available whenever users can livestream using the service)**	Confidential? – Y / N
Question 4: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 5: Do you have any views on the optimal design of reporting functions and choice categories for users to report content that depicts the risk of imminent physical harm ? Include any evidence, such as, testing to optimise wording, design of tools to support users to submit accurate and timely reports and how these may be used to support moderation actions.	Confidential? – N We note that the optimal design of any choice categories and reporting mechanisms will ultimately be highly service and context specific and determined by the service's existing choice architecture. We therefore recommend that design specifications are not included in any recommended measures but are left to the service provider. This will enable the service provider to ensure they implement proportionate measures, in accordance with the requirements of the OSA.

Question 6: Do you consider that there are alternative measures which would materially reduce the risks to users from livestreaming such as preventive safety by design frictions, prompts or restrictions? If so, please detail them and provide evidence on the costs and efficacy.	Confidential? – Y / N
Question 7: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Relevant proposal: ICU F3 (Providers should ensure that users are unable to do the following in relation to a one-to-many livestream by a child in the UK: a) Comment on the content of the livestream; b) Gift to the user broadcasting the livestream; c) React to the livestream; d) Use the service to screen capture or record the livestream; e) Where technically feasible, use other tools outside of the service to screen capture or record the livestream.)**	Confidential? – N **Measure ICU F3** We consider that Ofcom's requirements in relation to live-streaming are not sufficiently supported by evidence, and risk imposing disproportionate duties on service providers. *Insufficient consideration of the effect of the "first edition" safety measures* The obligations set out in the "first edition" Codes of Practice contained content moderation requirements that are likely to have an impact on both the type of content that is visible to a user during a livestream, and the user support and controls that a user has available. There has not yet been sufficient time to determine whether these new measures render these additional proposals obsolete: Ofcom should consider the effectiveness of these controls before recommending a raft of onerous new safety measures. Without considering the impact of these measures, Ofcom's evidence base for making recommendations is flawed. As safety measures dynamically come into place (for example, user control measures required by the first set of Codes) Ofcom should ensure that the available evidence from the first year of online safety compliance is better taken into account before proceeding further. *Limited consideration of the value of livestreaming* We set out our views on the economic and pro-growth value of livestreaming in response to Question 2. Ofcom should place greater weight on this evidence and consider the impacts of its measures on stifling growth and innovation. In addition, livestreaming can have social benefits. The restriction of comments or reactions in particular

	would greatly reduce the benefits of livestreaming, which we discuss further above. *Disproportionate approach* Under the OSA, services are required to implement proportionate systems and processes to protect children. This measure introduces blanket restrictions and the outright blocking of various functionalities for all children. This is disproportionate and unjustifiably extends Ofcom's previous approach in focusing on effective content moderation and the development of safeguards, while still allowing access to features. All services that offer one-to-many livestreaming and where it is possible for children to access part of the service are currently in scope of the draft measures. Ofcom could take a more risk-based approach that recognises only certain services with an elevated risk profile should be within scope of the draft measure. Introducing risk-based safeguards, as opposed to blanket restrictions, would be a more proportionate approach and we therefore encourage Ofcom to limit this measure to services that are at medium or high risk of grooming. We believe this is a more proportionate approach as it targets the measures at services that might be most at risk.
Question 8: If you are a service provider, what measures do you currently undertake to moderate livestreams and protect children who undertake livestream broadcasts, and what is your evidence on the effectiveness of such measures?	Confidential? – N We note that a number of Google's products currently adopt moderation methods that overlap with Ofcom's draft measures. For example, regardless of the age of the account holder, YouTube does not offer the ability to record or screen grab livestreams. Google's approach to livestreams has evolved over time recognising the specific risks on each service and we believe an approach that recognises different services may choose to adopt different moderation approaches would work more effectively than applying one-size-fits-all restrictions on livestreaming.

Question 9: Do you consider that there are alternative measures which would materially reduce the risks children face when livestreaming, both in general and in relation to operation of the supporting functionalities of comments, reactions, gifting and content capture? If so, please detail them and provide evidence on the costs and efficacy.	Confidential? – Y / N
Question 10: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 11: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Relevant proposals: ICU C11 (Assessing proactive technology for use to detect or support the detection of target illegal content.)** **PCU C9 (Assessing proactive technology for use to detect or support the detection of target content harmful to children.)** **ICU C12 (Assessing existing proactive technology for use to detect or support the detection of target illegal content.)** **PCU C10 (Assessing existing proactive technology for use to detect**	Confidential? – N **Measures regarding proactive technology** As currently drafted, the proactive technology measures appear to impose a general monitoring obligation on service providers which is both extremely difficult to implement, and directly contradictory to Parliament's intention. Under the Act, providers are required to take proportionate steps to mitigate the risk of harm. It is disproportionate to require services to implement these extensive requirements, not least because there is limited evidence on the efficacy of scanning in relation to certain types of content in particular, such as grooming, suicide content and eating disorder content. *Dramatic expansion of proactive technology measures* Measures ICU C11/12/13/14 and PCU C9/C10 substantially expand both the range of services in scope of proactive technology requirements and the harms for which service providers should proactively scan. We have significant concerns that in expanding the use of proactive technology to a host of further types of content,

or support the detection of target content harmful to children.)	Ofcom is applying excessively onerous duties that are not proportionate to the risks it has identified. We believe that such an expansion is disproportionate and that the breadth of the measure risks introducing a general monitoring obligation on service providers. As above, we note that Parliament's intent was explicitly not to include such an obligation. This measure is also not technically feasible in all cases, given it would require general monitoring of the service. This is particularly the case where Ofcom is applying proactive technology to content that is legal but harmful to children. In the absence of guiding law to determine what content falls into this category, providers may interpret this in different ways, which could lead to the over-removal of legal content that is not harmful to children (and in some cases may help them). For example, there are material difficulties involved in identifying some types of content harmful to children. What constitutes eating disorder content for one child may be recovery content for another, and there is a risk that requiring service providers to use proactive technology to prevent children from seeing this type of content may result in unnecessary over-removal and an associated negative impact on freedom of expression. The use of proactive technology must therefore be carefully balanced with the need to protect users' rights to privacy and freedom of expression. This was recognised during the passage of the Act, with clear constraints applied to its use, for example a recognition that proactive technology cannot be applied to private communications. We therefore request that Ofcom explicitly confirms (either in the Codes of Practice themselves or the accompanying policy statement) that none of the duties in the draft Codes would oblige platforms to proactively monitor the service for illegal and harmful content. We believe further clarity from Ofcom would add regulatory certainty and reduce the risk of significant over removal of legal content. *Effectiveness* Under the OSA, Ofcom is required to have regard to the degree of accuracy, effectiveness and lack of bias achieved by a technology in deciding whether to include it as a proactive technology measure in a Code of Practice (Schedule 4, para 13(6)). Ofcom has not demonstrated a compelling evidence base in support of this proposal.

	In particular, Ofcom has not shown that proactive technology is effective at identifying certain context dependent harms such as grooming, suicide and eating disorder content. Where there is a high amount of subjectivity in making a judgement about illegal or harmful content (for example, where the judgement relies in part on the intent of the person posting the content) or cultural nuance in understanding the meaning of a piece of content, proactive technology is less likely to accurately identify content that is harmful. Furthermore, technology aimed at proactively identifying violative content has to be developed at a global level. It is not feasible , proportionate or effective to build classifiers locally: this would mean developing classifiers aimed at proactively detecting unlawful content as defined by hundreds of laws in hundreds of jurisdictions. Instead, effective trust and safety work in this area is best carried out by developing technology that captures core types of harms that impact users globally. By mandating the nature and scope of these technologies, Ofcom is unintentionally making platforms choose either to develop less effective, UK-specific technology that will make users *less safe*, or apply UK standards to users worldwide, which is inappropriate and beyond the remit of the Act. *Targeting* We understand it is Ofcom's intent that service providers should only apply proactive technology to scan content related to the specific, in-scope kinds of illegal content for which a large service is at medium or high risk, or a non-large service is at high risk. However, as currently drafted, the measure appears to require any in-scope service to use proactive technology to detect: (a) in respect of the harm that the platform is medium / high risk for, content that amounts to an offence; and (b) all illegal content proxy on the service (where the terms of service prohibit the relevant harms), i.e. not restricted to the type of illegal content proxy that the service is medium / high risk for. We consider that (b) should be clarified to mirror the restriction in (a). It is disproportionate to require services to use proactive technology in relation to all illegal content proxy, rather than limiting it to the type of content for which a service is medium / high risk. We therefore suggest an amendment to the definition of "target illegal content" in ICU C11.5 to ensure that proactive scanning only relates to specific illegal content:

	"target illegal content" means relevant content ***in relation to a relevant harm*** *that either: a) amounts to an offence; or b) is illegal content proxy, where the provider is satisfied that its terms of service prohibit the relevant harms.* *Costs of proactive monitoring* Paragraph 9.116 of Ofcom's policy statement acknowledges that the ongoing maintenance cost of proactive technologies can be tens of millions of dollars per year for large and complex services. Ofcom do provide indicative figures for third party scanning solutions but Ofcom should reflect industry evidence and recognise that costs across large services are likely to represent significantly more than currently estimated. Given the extremely high costs in relation to this type of technology, Ofcom should consider whether the recommended measures remain proportionate. *Regularity of proactive technology assessment* In respect of a number of proactive technology measures (e.g. ICU C11, C12, C13), there is a requirement to assess the tech at least every 12 months or earlier where there is a significant change to the service (in line with updates to the risk assessment). It is unduly burdensome to require services to review the technology this frequently, given that it is not clear that a significant change to the service and its risk profile would necessarily be correlated with the ability to deploy proactive technology. For example, a change to content prioritisation rules on a service may constitute a significant change for risk assessment purposes but is unlikely to affect a service's ability to use proactive technology. At the very least, rather than requiring an annual assessment in addition to an assessment after a significant change, providers should have more discretion about whether the significant change is one that could justify a re-assessment of the technology. If Ofcom decide to proceed with introducing these measures, we therefore recommend that they are only required to reassess the proactive technology every 12 months or where they consider it necessary in light of changes to the service. *Technological Infeasibility* Service providers benefit from the harbour of the draft Codes and need not deploy proactive technology where they judge it is technically infeasible to implement proactive technology on their service. Ofcom

	should explicitly recognise that service providers need not apply proactive technology on E2EE content. Without this explicit recognition, service providers will continue to be under an obligation to conduct an annual proactive technology assessment and operate with a lack of regulatory certainty about Ofcom's view about the applicability of proactive technologies in E2EE environments.
Question 12: Do you have any comments on the Proactive Technology Draft Guidance?	Confidential? – Y / N Google is constantly improving and introducing changes to support online safety and tackle illegal and harmful content at scale, including through the use of automated content moderation. We appreciate that Ofcom is trying to set clear measures that are accomplishable for a range of services. However, in general we consider that prescriptive rules inhibit our ability to prioritise safety according to the trends we see on our platforms and the risks we identify in our risk assessments. We consider there is a risk that prescriptive provisions that fail to address the most pressing issues (some of which could be unique to a service and therefore not able to be identified in general guidance), could unintentionally push platforms to adopt a "lowest common denominator" approach to comply with the law, rather than take a more sophisticated and effective approach that would not offer the same certainty of compliance. We would welcome it if the guidance could be framed more broadly in places, to ensure that a range of current and future technological solutions to compliance can benefit from the safe harbour provisions.
Question 13: Do you agree with the harms currently in scope of these measures? Are there any additional harms that these measures should capture? Please provide the underlying arguments and evidence that support your views, including evidence regarding the availability of accurate and effective proactive technology.	Confidential? – Y/N

Question 14: Do you agree with who we propose should implement these measures? Are there any other services that should be captured for some or all of the relevant harms?	Confidential? – Y / N
Question 15: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 16: Do you agree with our proposal? Please provide your reasoning, and if possible, provide supporting evidence.	Confidential? – N *Technical Feasibility* The ICJG recognises that it may not always be technically feasible to review message content itself or connected content. We would welcome explicit confirmation that where content or connected content is E2EE it would be technically infeasible to assess this content. *Impact on users* We note that there are significant consequences for users where service providers identify CSAM, including law enforcement reporting and user sanctions (including a proposal to ban users from the service and prevent their return). The proposed amendments to the ICJG would allow service providers to infer CSAM content on the basis of very limited evidence, where it is not possible to review the specific content within message forums. There is therefore a risk that service providers impose significant sanctions erroneously on a user without correctly identifying CSAM. This risk is especially acute where service providers have only limited information, broader evidence is not reasonably available and there is no evidence to the contrary. *Proportionality* As drafted, the ICJG suggests that service providers may have to consider several sources of reasonably available information to make a judgement of illegality. What constitutes “reasonably available” is

	unclear and this would represent a disproportionate burden were services to be required to actively seek out this information and make an assessment having considered several sources.
Question 17: Do you have any evidence relevant to the examples given?	Confidential? – Y / N
Question 18: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 19: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Relevant proposals: ICU C14 Providers use perceptual hash matching to detect image-based intimate image abuse content so it can be removed.** **ICS C8 Providers use perceptual hash matching to detect image-based intimate image abuse content so it can be moderated.**	Confidential? – **YES** ()
Question 20: Do you have any evidence on the relative efficacy of third-party and internal databases for image-based IIA content?	Confidential? – Y / N

Question 21: Do you consider this measure to be effective for file-sharing and file-storage services? Please explain your reasoning and, if possible, provide supporting evidence.	Confidential? – Y / N
Question 22: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 23: Do you consider this measure to be effective for large general search services? Please explain your reasoning and, if possible, provide supporting evidence. **Relevant proposal: Providers use perceptual hash matching to detect image-based intimate image abuse content so it can be moderated.**	**Confidential? – Yes** (□)
Question 24: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.	Confidential? – Y / N
Question 25: Do you have evidence regarding the accuracy and effectiveness of hash matching solutions for detection of terrorism content specifically (including their	Confidential? – N Google adopts a flexible and multidimensional approach to the detection and prevention of the proliferation of terrorism content across its services, relying on both automated and human efforts to identify potentially violative content. Hash-matching is just one of the automated methods employed, alongside machine learning classifiers.

false positive and false negative rates)? **Relevant proposal: ICU C13 Providers use perceptual hash matching to detect terrorism content so that it can be removed.**	However, as detailed above in Question 23, we have found that hash-matching often proves to be an ineffective way of uncovering certain types of violative content. Google offers numerous products with varying functions. On our user-to-user services that host user-generated content (such as Google Drive, YouTube and Google Photos) we are, in some cases, able to use hash-matching to identify terrorism content by checking content that has been shared against a repository of known violent extremist content. However, more generally, the nature of terrorism content contributes to why we have found hash-matching often proves less effective for this type of content, in practice. Terrorist and violent extremist content can appear on each product differently and consequently, we adopt slightly different policies across each product. Likewise, whether material or content can be correctly classified as "terrorism content' depends on its context and purpose. While hash-matching is effective in detecting a particular known image, it is much less effective in determining the context of that image or the purpose for which a user may have shared that content. There are many legitimate, non-malicious and lawful reasons why material that may appear as terrorism content may be shared or uploaded on Google's platforms. For example, in one context, footage of a terrorist attack might be documentary evidence of atrocities in areas that journalists have great difficulty accessing. However, in another context, the footage could be promotional material for a terrorist organisation. Where context is important in determining an accurate classification of content, the use of perceptual hash-matching proves less effective and it is preferable to use a combination of cryptographic hashing and other content moderation processes. Taking into consideration the volume of material stored on services such as Google Drive and the variety of purposes one may use a Drive account for, extensive secondary human review is required to try to ascertain the relevant context of the content. For example, where Google does deploy automated tools, 96% of the unique items flagged require human review. Google adopts alternative methods in circumstances where hash-matching proves less effective. This includes user reporting via

	in-product tools, engaging with external partners and training our other automated tools to quickly identify violative content. The evidence therefore does not support the proposal that mandating the use of perceptual hash-matching will ensure or improve the detection of terrorism content. Consequently, Google chooses not to use specific types of technology in isolation, in order to ensure that the overall package of systems deployed is sufficiently precise. Should hash-matching be adopted as a recommended measure, it is noted that, as with IIA, Google receives hashes of terrorism content in cryptographic form rather than in perceptual form. As advised above, we request that Ofcom does not mandate the use of *perceptual* hash matching and instead leaves room for choice as regards the form of hash matching adopted. We therefore reiterate the recommendation made earlier in our response that the most effective approach for Ofcom to take is to give service providers flexibility in relation to the types of technologies that are most likely to be effective at identifying terrorism content on their platforms.
Question 26: Do you have evidence on the extent to which a hash matching solution can identify terrorism content accurately when applied in different contexts from that in which the hash was created, noting the potential implications for freedom of expression;	Confidential? – Y / N
Question 27: Do you have a view on the degree of human oversight required to support the use of hash matching in relation to terrorism content?	Confidential? – Y / N

Question 28: Do you have evidence or views on the impact assessment (including costs) associated with implementing and maintaining hash matching technology for the detection of terrorism content (such as the impacts and costs of setting up an internal database, connecting to an external provider, and moderation costs).	Confidential? – Y / N
Question 29: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Amendment to ICU C9** **Providers should ensure that hash-matching technology is used to detect and remove child sexual abuse material (CSAM).**	Confidential? – Y / N
Question 30: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 31: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Relevant proposal: ICU E2 Providers should design and operate their recommender systems to ensure that content indicated potentially to be certain kinds of priority illegal content (relevant illegal content) is excluded**	Confidential? – N **Measure ICU E2** We recommend that this measure is removed from the Code of Practice, as it is not aligned with the scope of the OSA, is unlikely to be effective and risks encroaching on users' freedom of expression. It is also clearly inconsistent with the current Codes of Practise (ICU C1 and ICU C2) which require "*when the provider determines that: a) the content is illegal content (pursuant to ICU C1.3(a))...the provider should swiftly take the content down*". This existing requirement is aligned with the

from the recommender feeds of users.	Act, and ensures users' rights to freedom of expression. It requires platforms to 'review and action' rather than 'guess and action'. While the safety goals are clear and vital, there is a delicate balance to strike which the measure as currently drafted does not achieve. The proposals introduce risks of false positives, where legitimate content, including important political or journalistic speech, might be unfairly excluded from recommendations. Such exclusions could also increase the workload for content review teams and potentially erode user trust if people feel their control over their experience is diminished or if the system frequently makes mistakes. *Disproportionate expansion of the type of content regulated by the OSA* In passing the OSA, Parliament made a deliberate legislative choice to confine the safety duties on service providers to illegal content, content harmful to children, fraudulent advertisements and pornographic content. Each of these categories of content are defined in detail in the Act. In doing so, Parliament decided that the OSA would **not** extend to content that was lawful but potentially harmful to adults. This approach was debated extensively in Parliament, and the legislative intent that this should not be in scope of the duties under the Act was clear. The requirement on service providers to identify "potentially" illegal content is therefore at odds with the statute. This is because this approach, in effect, "upgrades" this kind of content to the status of illegal content. *Disproportionate expansion of the type of judgements being made by a provider* Relatedly, Parliament also specified the approach to be taken where a system or process of a provider involves a judgement about whether content is content of a particular kind, and in particular about whether content is illegal content (s.192 OSA). It is therefore clear on the face of the Act how providers should identify content in scope of the Act. Contrary to this, measure ICU E2 appears to introduce a new category of "potentially" illegal content for service providers to identify that was neither foreseen or intended by the Act or intended by Parliament to be captured, nor is there any guidance on how to identify this type of content. This measure consequently risks introducing an unduly burdensome requirement on service providers.

	The proposal therefore appears to amount to a general monitoring obligation which is both extremely difficult to implement, and directly contradictory to Parliament's intention. From a technological point of view, it would be necessary to automatically monitor, assess and evaluate all content on the platform and compare it against a nebulous category of "potentially" illegal content that is undefined in law. *Interference with users' freedom of expression* There is a significant risk of an adverse impact on freedom of expression from over-removal in relation to this novel, broad category of content and as a result of the requirement to exclude this type of content from recommender feeds. This specific requirement has far reaching implications on access to information. It is clear that the temporary exclusion of lawful content also represents an interference in users' freedom of expression. This interference will be amplified where service providers have no guidance on how to judge which content is potentially illegal. In the absence of this, service providers are at risk of making inaccurate decisions and "over-remove" content that is in fact legal. *Limited effectiveness* As set out in the measure, Ofcom clearly acknowledges that content excluded from recommendations might later be "determined not to be illegal content". The proposal also acknowledges that there is "limited evidence available" on the accuracy of technology for identifying "potentially illegal" content. It is unacceptable for Ofcom to proceed with this measure given this limited evidence base. Erroneous flagging, especially for "highly protected forms of speech such as religious expression... or political speech; or... journalistic content," creates a potential interference with freedom of expression. This measure also underestimates the way in which alternative approaches that do not risk curtailing users' freedom of expression can be used to promote relevant, safe content. In particular, the use of personalization helps users find content that inspires, informs and entertains them, and plays a crucial role in maintaining a responsible platform. An unfocused requirement to change how content recommender systems work on the basis of potential illegality could negatively impact the user experience as lawful content is removed. This could lead to more irrelevant or and lower-quality videos being

	promoted to users, worsening user experience for no appreciable safety benefit. Furthermore, service providers have already invested significant time in determining what kinds of content should be prohibited in their terms of service or community guidelines. It may be the case that "borderline" content is already prohibited. Service providers may therefore already remove illegal content proxy as part of the existing safety duties set out in the first edition Codes. This content will have been clearly defined by service providers. It is also clear to users the basis on which their content is being removed. The addition of a further indefinite category of "potentially" illegal content therefore does not contribute to keeping users safe. *Poorly targeted* The measure applies to all services with a content recommender system that are medium or high risk of specific harms (terrorism; hate; encouraging or assisting suicide; foreign interference). As set out in our response to the protection of children Code of Practice, we encourage Ofcom to reconsider how it positions recommender systems as largely a negative risk factor. Recommendations play an important role in how services maintain a responsible platform, including by connecting young users with age appropriate content, in addition to helping connect viewers with content that uniquely inspires, informs and entertains them. Ofcom's conclusion about harm associated with recommender systems does not take into account the positive effects that a recommender system may have. Recommender systems complement the work we do to remove content that violates our Community Guidelines or the law in the countries where we operate, such as the UK. The measure is poorly targeted as even services which operate recommender systems that do not promote quick and wide dissemination of content but are, for example, merely ancillary or used for moderation or safety purposes are included. Ofcom should seek to understand which content recommender systems may actually increase the risk of harmful content were it minded to pursue this measure, and to specify that it is only in respect of those content recommender systems that the measure applies. It is further poorly targeted as the measure requires that services at medium or high risk of a specific harm must apply the measure to all

	categories of content. This would require significant changes to how recommender systems operate. For example, the signals that may indicate potentially illegal fraudulent content are likely to be different to those that may indicate potentially illegal terrorism content. This measure could require services to change their recommender system to identify potentially illegal content for those harms where they have low/negligible risk. We believe that such an approach is disproportionate and the measure could be more effectively targeted. *Disproportionate burden on service providers* The proposal mandates excluding content based on indicators of being potentially illegal, unless and until content moderation teams have reviewed it. If there is an increase in videos flagged as potentially illegal, without an increase in reviewer capability, any sort of expectation of short, fixed turnaround times would compromise accuracy and risk creating unnecessary censorship. Services are therefore required to balance the accuracy of review with the turnaround time. Expanding the scope of content requiring human review to include "potentially illegal content" could significantly increase the workload for review teams, potentially leading to backlogs and delays in content assessment, including for lawful content. This would hinder the stated aim of addressing "incoming appeal requests as quickly as possible", and risks legal content being downranked or removed from the platform unnecessarily. This has the potential to significantly impact users' freedom of expression. *International Divergence* The proposal as drafted would require service providers to moderate legal, non-harmful content for both adults and children. Such an approach would be out of step with other international regulatory regimes (the DSA does not envisage action on lawful content for example).
Question 32: Do you have evidence on what types of content are typically recommended to users as part of concerted foreign interference activity;	Confidential? – N

Question 33: Do you have evidence on whether services track the extent of algorithmic amplification, such as impressions and reach, of content that is later deemed illegal/violating. If so, do they (or does your service) use this information to enhance the safety of their systems?	Confidential? – Y / N
Question 34: Do you agree with our assessment of the impacts (including costs) associated with this proposal? Please provide any relevant evidence which supports your position.	Confidential? – N As above, we consider that it is not proportionate for Ofcom to regulate the visibility of lawful content, as this represents an interference with a user's freedom of expression.
Question 35: Are there any impacts of the proposed measure that we have not identified? Please provide the rationale and any supporting evidence for your response.	Confidential? – Y / N
Question 36: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Relevant proposal: ICU H3 Providers should ban users who share, generate, or upload CSEA, and those who receive CSAM, and take steps to prevent their return to the service for the duration of the ban**	Confidential? – N We are aligned with Ofcom's intention to impose serious consequences on perpetrators of CSAM offences and support some key elements of the proposals. In particular, we agree with the provisions in ICU H3.4-3.6 that allow providers, to some extent, to consider what is reasonable in imposing measures against perpetrators. There are some elements of this measure, however, which risk causing unintended consequences, and place a disproportionate burden on service providers who are not well equipped to make legal judgments. <u>*User banning for receipt of CSAM*</u> We note that it is an offence for an individual to possess CSAM, and we assume that this is the reason that measure ICU H3 extends the ban on users to those who "receive" CSAM. However, defences to this offence may apply where i) the recipient did not see the content and had no

	reason to suspect it was CSAM or ii) the content was received without prior request and not kept for an unreasonable time. Where these defences apply, a user has not committed an offence and need not be banned for receiving CSAM. Ofcom states (footnote 483 in the policy statement) that a service needs reasonable grounds to infer that the defence applies to judge that the receipt was not illegal and take no action against the recipient. It is extremely difficult, if not impossible, for service providers to have positive grounds to infer that the defence does apply without carrying out disproportionately extensive further inquiries. This is in part because such a finding would rely on a judgement of the state of mind of the recipient or knowledge that no request was made to receive an image (which may well have happened on a separate service, or in separate messages / communications that the provider has not reviewed). We therefore anticipate that recipients of CSAM will be banned under the proposed measure and this may lead to perverse outcomes. For example, in some scenarios this may lead to the perverse incentive for users to share CSAM with multiple users with the intention of getting them banned from several services. We therefore question if this measure will have its intended effect of reducing the circulation of CSAM online. Furthermore, it may reduce the incentive for users to report CSAM, because the fact that they have encountered the CSAM such that they are able to report it likely means that, under the current terms of the measure, they should be banned from the service. This point was raised by the NSPCC during a House of Lords Communications and Digital oral evidence session. Rani Govender, policy and regulatory manager for child safety online, stated: "*If we just leap into bans, that scares children away. We want to make sure that we are never deterring children from speaking up about something that has happened to them, whether that is something they have experienced or something they might have done online*" (transcript here). We therefore recommend that Ofcom restrict the application of this measure to users who share, generate or upload CSEA. If Ofcom decides to proceed with this measure as currently drafted, we would welcome further practical guidance of where defences may apply and how service providers should judge this, in particular if whether a user report is indication that one of the defences applies.

	ICU H3.6 - child users sending/receiving self-generated indecent imagery We support Ofcom's draft proposal at ICU H3.6 to give service providers the discretion not to ban child users where the provider has reasonable grounds to infer that the child user has i) shared self-generated indecent imagery (SGII) as a result of being groomed or ii) shared or received SGII as part of an age-appropriate relationship. The proposal recognises that users should not be banned for sharing indecent imagery where they are either a victim or exchanging imagery as part of an age-appropriate relationship (which Ofcom recognises is an increasingly common part of older children's relationships). We would welcome further clarity on the types of information a service may use to determine a user is a child in the absence of that user having completed an age verification process (paragraph 16.71). *Continuous, cross-service bans requiring steps to prevent a user from regaining access* Firstly, as set out above, we are aligned with Ofcom's intention to impose serious consequences on perpetrators of CSAM offences. However, we consider that this measure could benefit from further flexibility to allow service providers to carry out a contextual review of a user's account. In particular, there may be circumstances where the context is necessary to understand why the image has been shared or received. For example, a police officer/social worker that has CSAM in their account due to an accidental upload or poor data hygiene, or someone engaging in misguided activism. We welcome Ofcom's recognition at ICU H3.4 of the draft Codes that *"What "reasonable steps" amount to in this context will depend on, among other things, the nature of the service in question, what is technically feasible, the resources of the provider, and the provider's risk profile, having regard to the provider's risk assessment*." We would welcome Ofcom's additional clarification that service providers may consider the effectiveness of any measures in a judgement on what is reasonable. This would be a more proportionate approach to ensuring service providers implement effective solutions that stop offenders returning to services. *Duration of bans*

	We agree with Ofcom's proposed Option 2 set out in its policy statement (para 16.33) where *"service providers are given discretion as to what would be an appropriate length of a ban"*. In general we note that service providers will be supportive of lengthy, or indeed permanent, bans of users who share or upload CSAM given the severity and duration of the harm that such content can cause to victims and those that encounter such content. However, we believe that service providers should be afforded discretion in determining the length of bans. As above, there may be circumstances in which a service has no positive grounds to infer that a defence applies. For example, where there is circumstantial evidence that a user did not want to receive CSAM. For example, there may be a scenario in which a single user shares CSAM content with multiple unconnected users who did not wish to receive the content (but a service provider has no positive grounds to infer a defence applies). In this scenario, we believe that there would still be a legitimate reason to enforce a less severe ban on those users who had received the CSAM content (and we note that this circumstance would not fall within Ofcom's definition of an "exceptional circumstance" at ICU D10. 5 were Option 3 to be pursued). For certain service providers who operate multiple services, enforcing such a ban would represent a significant interference in an individual's day-to-day life. The above scenario is just one example and we note there may be multiple, similar "edge" cases. Given the potential for similarly difficult judgements, we would caution Ofcom against such a prescriptive approach and support greater service provider discretion. In addition to the difficulty of making subjective judgements, we would support the evidence provided to the House of Lords Communication and Digital Committee. Attendees suggested that service providers should take a more tolerant and educative approach when sanctioning child users involved in sharing or receiving CSAM (here). Attendees raised the fact that permanently banning children was not in the public interest as it represented "punishment upon punishment", may reduce the likelihood that children speak up about CSAM harms online and it would exclude child users from important digital spaces. While providing service providers with discretion, we think the objective of ensuring that genuine bad actors receive lengthy bans could be achieved via Ofcom engagement and guidance to service providers

	(and as above, Google is fully supportive of taking robust action against bad actors who share CSAM content with genuine intent). In the event Option 2 is not pursued, we agree with Ofcom adopting Option 3 and service providers having the option of not enacting permanent bans in exceptional circumstances.
Question 37: What is your assessment of the options we set out in relation to the treatment of child users and which option do you consider to be most appropriate? Please provide any supporting evidence to support your arguments.	Confidential? – Y / N
Question 38: Do you agree with our assessment of the impacts (including costs) associated with this proposal? Please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 39: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Relevant proposal: ICU H2 Providers should prepare and apply a sanctions policy in respect of UK users who generate, upload, or share illegal content and/or illegal content proxy / content harmful to children, with the objective of preventing future dissemination of illegal content / CHC.**	Confidential? – Y / N

Question 40: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 41: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **ICU B1 A measure that defines highly effective age assurance for the purposes of the Illegal Content User-to-user Codes and sets out principles that providers should have regard to when implementing an age assurance process.** **ICU D15 Providers should allow for appeals of highly effective age assurance decisions and take appropriate action where these are upheld.** **ICU D16 Providers should allow for appeals of highly effective age assurance decisions and take appropriate action where these are upheld.** **PCU B1 Amendments to codify the definition of highly effective age assurance in the Protection of Children User-to-user Code**	Confidential? – Y / N

Question 42: Do you agree with our proposal to introduce age assessment appeals measures into the Illegal Content User-to-user Codes (ICU D15 and D16)? Please explain your reasoning.	Confidential? – Y / N
Question 43: Do you agree with our proposed amendments to codify the definition of highly effective age assurance in the Protection of Children User-to-user Code? Please explain your reasoning.	Confidential? – Y / N
Question 44: Do you agree with our proposed amendments to the Part 3 Highly Effective Age Assurance Guidance? Please explain your reasoning.	Confidential? – Y / N
Question 45: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 46: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **ICU F1 and F2** **We are recommending amendments to ICU F1 and F2. These proposed changes would mean that services in scope of ICU F1 and F2 should implement these measures using one of the following approaches: • Using**	Confidential? – Y / N

highly effective age assurance, as set out in the Part 3 HEAA Guidance; or • Applying ICU F1 and F2 to all users of the service	
Question 47: Do you agree with option A and option B in increasing the effectiveness of the ICU F1 and F2 measures?	Confidential? – Y / N
Question 48: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 49: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence. **Relevant proposal: ICU C15 / PCU C 11 The provider should prepare and apply an internal crisis response protocol. It should also conduct and record a postcrisis analysis. Providers of large services should implement a dedicated communication channel by which law enforcement can contact them on crisis-related matters during a cri**	Confidential? – N **Post-crisis analysis** We would recommend that Ofcom add a requirement that post-crisis analysis need only be conducted where the service provider has reason to believe that lessons from the analysis may usefully inform the future development of the internal crisis protocol. At present, the measure may impose an unnecessary administrative burden where service providers are conducting in-depth analysis without learning anything new about content trends and their crisis response mechanisms. For example, if the crisis is unique and has not been seen before by the service provider, it may merit such analysis and could usefully be fed into an updated version of the crisis protocol. However, if the crisis is very similar to one experienced previously, this may not be necessary. **General approach to crisis protocols** Generally, we consider that Ofcom's analysis of the costs of preparing a crisis protocol appear arbitrary. In particular, the requirement to re-allocate content moderators is disproportionate, and service providers should be allowed flexibility to assign resources as they see fit, given the global nature of content moderation and the global

	challenges the service faces on any given day. We therefore recommend that Ofcom recognise this in their policy statement. Furthermore, we recommend that Ofcom explicitly states that service providers may have a combined crisis response plan, rather than two separate plans (one in respect of illegal harms and another in respect of content harmful to children).
Question 50: Do you agree with our proposed definition of 'crisis'? Please explain your reasoning, and if possible, provide supporting evidence.	Confidential? – Y / N
Question 51: Do you consider these measures to be effective for services that are not large services? Please provide any evidence on the role of services that are not large services during crises.	Confidential? – Y / N
Question 52: Is there any evidence of best practice in responding to a crisis that we have not identified? Please explain your reasoning, and if possible, provide supporting evidence.	Confidential? – Y / N
Question 53: Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 54: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.	Confidential? – N Ofcom's proposed amendments would require service providers to extend the ability to appeal actions taken by the providers to cover action in relation to 'proxies' for illegal content and content harmful to

Various across ICU D and PCU D Broadening the scope of appeals measures to ensure that they cover decisions taken on the basis that content was an 'illegal content proxy	children. This means that users would have the ability to appeal decisions taken by a service provider not only where content is actually deemed to be illegal or harmful to children, but also in violation of policies that are "proxies" for such content - that is, under ambiguous circumstances where the provider merely "has reason to suspect" that the content "may be" illegal or harmful to children. Under the OSA, Category 1 service providers are required to operate complaints procedures that allow users to make appeals in respect of content that violates their terms of service, that is, illegal content or harmful-to-children proxy (s.72(6)-(8) OSA). The decision to limit this duty to Category 1 services demonstrates Parliament's clear intention to impose this obligation on only this smaller group of services, as opposed to all regulated services more broadly. Ofcom's proposed amendment to the measure would have the effect of introducing this requirement on all regulated services by the back door, contrary to Parliament's intention. Given the very broad and ambiguous definition of illegal content proxy there is even more reason to limit this to Category 1. This amendment should therefore be removed from the final version of the Codes.
Question 55: Do you agree with our assessment of the impacts (including costs) associated with this proposal? Please provide any relevant evidence which supports your position.	Confidential? – Y / N
Question 56: Do you think our package of proposed measures is proportionate for services in scope of the Illegal Content User-to-User Codes, taking into account the existing package of measures, the impact on reducing the risk of relevant harms and the implications on different kinds of services?	Confidential? – Y / N

Question 57: Do you think our package of proposed measures is proportionate for services in scope of the Protection of Children User-to-User Code, taking into account the existing package of measures, the impact on reducing the risk of relevant harms and the implications on different kinds of services?	Confidential? – Y / N
Question 58: In relation to our equality impact assessment, do you agree that some of our proposals would have a positive impact on certain groups? Please explain your reasoning and provide supporting evidence where possible.	Confidential? – Y / N
Question 59: Do you consider that our proposals could have any negative impacts on certain groups? If so, please explain your reasoning.	Confidential? – Y / N
Question 60: In relation to our Welsh language assessment, do you agree that our proposals are likely to have positive, or more positive impacts on opportunities to use Welsh and treating Welsh no less favourably than English? If you disagree, please explain why, including how you consider these proposals could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.	Confidential? – Y / N