Consultation response form

Please complete this form in full and return to protectingchildren@ofcom.org.uk.

Consultation title	Consultation: Protecting children from harms online
Organisation name	Google

Summary

We look forward to working closely with Ofcom to help inform a regulatory approach to child safety which protects children from harmful content, while preserving the core benefits of online environments, including the ability of children and adults to express themselves openly, access useful information and make positive connections.

At Google, we spend a lot of time developing policies to protect our youngest users and take seriously the heightened social responsibility to help protect children and teens while ensuring that they have access to healthy and helpful online experiences.

We support Ofcom's efforts to implement a regime that requires online services to build age-appropriate products; offer tools that give families flexibility to manage their relationships with technology; and implement policies, protections, and programs that increase child safety online. We believe that the comments and suggestions provided in this response are a constructive and useful contribution to inform Ofcom's thinking on this vitally important aspect of the Online Safety Act regime.

We are broadly aligned with Ofcom's approach in the draft Codes and recognise the challenges of drafting measures that must be interpreted and applied across a wide variety of services. We have a number of practical suggestions which we consider will improve the workability and proportionality of the draft Codes consistently with the policy intention of the Act.

We have raised concerns about areas where the Codes become too prescriptive or disproportionate, requiring services to make changes that may not materially improve the safety of younger users, and could potentially distract from our shared core objectives to keep children safer online. We are keen to ensure that the obligations are not too narrowly defined, allow for flexibility, and do not unintentionally stop adults and children accessing vital online services due to overly burdensome, ineffective or inappropriate age assurance measures.

In summary:

<u>**Age Assurance**</u>

Age Assurance must be risk-based: We agree with the risk based framework for age assurance that was set out in the Online Safety Act and believe that it could be better reflected in the Codes. While age assurance is an important tool, we think it should be used in a way that is risk-based and proportionate to the level of potential harm on a particular service. It should also be one of a range of tools used by services to protect children.

We think it is important that the Code allows providers of user-to-user services to use age inference models to comply with the child safety duties where appropriate, as is consistent with the Act. Services should not be forced to assume a user is a child where an age inference model indicates that they are above 18, as this risks resulting in an inappropriate curtailment of the rights of adult users with a profound societal and economic impact on such users. **Where age inference models are**

sufficiently accurate and used in conjunction with other safety tools, they should therefore be considered to be an acceptable form of highly effective age assurance. Where these models are used, it would be appropriate for Ofcom to require the model to operate within a reasonable level of accuracy, as opposed to imposing a specific accuracy range, which may be disproportionate given the risk that the service poses to users. If a service has reasonable grounds to conclude, based on age inference, that a user is an adult, services should be able to treat this user as such.

We consider that the approach to age assurance taken by Ofcom must be consistent with the need to carry out no more than the minimum level of processing of user's data. We also consider that a proportionate approach would allow less intrusive age assurance methods to be used on lower risk services, as we explain in more detail below.

Age assurance methods need to be tailored to the particular risks on a service rather than using a one-size-fits all approach. As currently drafted, the Codes and Ofcom's guidance do not appear to acknowledge that different types of content and services require different types of age assurance (and the term "highly effective" should be interpreted accordingly). While maintaining robust standards that ensure children are kept safe online, **Ofcom should look at a number of factors when considering what is "highly effective": for example, this could be different according to the risk level of a particular service, the nature of the content on that service, and/or how the service is accessed.** We believe this flexibility is needed because (as reflected in the ICO's Children's Code) the "harder" the method of age assurance, the more intrusive and data intensive the method and the higher the impact on all users' ability to access services and information. In particular, more intrusive age assurance methods could have unintended consequences such as locking adult users out of vital online services and causing a negative economic impact on a variety of legitimate services. Methods for establishing the age of the user therefore need to be inclusive and take into account the very diverse set of users that we serve. Not all users have access to credit cards or have government issued IDs. Furthermore, individuals from marginalised or vulnerable groups may be unable or unwilling to share this information, further compounding their disadvantage. As noted in Google's Legislative Framework to Protect Children and Teens Online, we think that the most data intrusive methods (such as verification with "hard identifiers" like government IDs) should be limited to high-risk services (i.e., those that are high risk for primary priority content) or age correction.

Similarly, the Codes 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. The use of "highly effective" age assurance is required in both cases (and, as above, we understand that Ofcom interprets this term uniformly across different services with different levels of risk). Not only is this disproportionate in applying the same burden to services that pose different levels of risk (with a commensurate negative impact on adult users of those services) but it also removes an incentive for high risk services to reduce risk. **It would be more proportionate for "highly effective" to be interpreted in a context-based way such that different types of age assurance are accepted as highly effective, depending on the level of risk on a service.** This would be consistent with paragraph 12(2)(d) of Schedule 4 of the Act, which requires age assurance measures in the codes of practice to reflect "the principle that more effective kinds of age assurance should be used to deal with higher levels of risk of harm to children".

As Ofcom recognises in the draft Code, age assurance is not a silver bullet and must be used alongside other tools. **Age assurance should be considered alongside a variety of other tools that platforms can use to keep children safer.** For a platform such as YouTube, this would include reporting mechanisms for harmful content, curated spaces for children (such as YouTube Kids), recommender systems which promote helpful, age appropriate and authoritative content, and tools for parents. **We are concerned that Ofcom's Code places too much emphasis on age assurance without providing flexibility for lower risk services to meet the objective of protecting children by other means.**

As an example, parental controls such as Google's Family Link enable parents to better control their child's online experiences (e.g. only parents can turn Safe Search settings off for their child). It can also provide a useful age assurance tool in itself by allowing parents to confirm the age of a child, which we consider should be regarded as highly effective age assurance in Ofcom's guidance.

Recommender Systems

We would encourage Ofcom to reconsider how it positions recommender systems as largely a negative risk vector. 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.

On YouTube, recommendations complement the work we do to remove content that violates our Community Guidelines (including Primary Priority and Priority Content that is harmful to children). More specifically, when it comes to our younger users on YouTube, recommendations help connect them to high-quality content and minimise the chances they'll see low quality content. For example, we raise high-quality children and families content in recommendations that meet our quality principles. This is content that inspires curiosity, imagination, and creativity in younger viewers while helping older children consider a diverse range of cultural and societal issues which they may not be exposed to. These high-quality content principles were developed in collaboration with our Youth and Families Advisory Committee, which consists of 13 global experts in children's media, child development, digital learning, and citizenship from a range of academic, non-profit, and clinical backgrounds. We consulted with the same experts to develop a companion set of low-quality content principles that impact channel performance. We use recommendations to reduce the spread of content that matches our low-quality principles — for instance, because it's heavily commercial or pseudo-educational.

In addition to recognizing the benefits of recommender systems for younger users, **the Code's requirements on recommender systems should explicitly recognise that they do not apply to content that a user has previously encountered, or which is already within a user's own account, since the policy intent is to ensure that children are not recommended harmful third-party content** (see, for example, paragraph 3.19 of Ofcom's advice to Government on the categorisation of services). For example, Memories recommended by Google Photos from a user's own library should not fall within scope.

We note that the various measures proposed in relation to recommender systems risks effectively requiring platforms to carry out general monitoring, given that they can only be met by proactively reviewing the majority of, if not all, content on a platform for the purpose of identifying a huge variety of broadly-defined harms. This was clearly not the intention of the Government or Parliament (as per Lord Parkinson who confirmed that the Act "does not require general monitoring of all content"- Session 1, 27 April 2023), and we do not consider it to be consistent with the privacy and freedom of expression safeguards in the Act. In particular, we note the difficulties of scanning for Ofcom's proposed categories of non-designated content, which we comment on in more detail at Section 7 below.

There is a significant risk of an adverse impact on freedom of expression from over-removal where content categories are defined broadly and as a result of the requirement to give lower prominence to any other piece of content that 'shares significant characteristics' with content towards which a child user has signalled negative sentiment. This specific requirement has far reaching implications on access to information.

Providing sufficient flexibility

As with the Illegal Harms Consultation, we are concerned that the more prescriptive measures in the Code do not provide sufficient flexibility to allow services that are experienced in online safety to use the most effective methods for mitigating risk.

For example, we note the measure requiring services to put in place a prioritisation framework for content review based on virality, potential severity of harm and the likelihood of content being content that is harmful to children. **In order to give services sufficient flexibility to retain prioritisation processes that have been carefully developed and are demonstrably efficient and robust, we suggest that Ofcom instead provides examples of factors that should be taken into account**. It would be more proportionate for the Code to set an objective (e.g. a requirement to ensure that content most likely to be harmful to child users is prioritised for review or downranked) with examples of factors, such as virality and potential severity of harm, provided as those that **may** be taken into account. Framing the Codes more broadly in this way would help ensure that a range of current and future technological solutions to compliance can benefit from the Code's safe harbour, while allowing us to continually innovate in online safety, leverage improvements in AI to improve user safety; and use processes that are technically feasible and proportionate to execute at the scale of Google's operations.

Blocking/muting obligations

As set out in the draft Illegal Harms Codes, (see the answer to Question 37 (20.1)), the detailed requirements set out by Ofcom in relation to blocking and muting users are too prescriptive and disproportionate as they would would not reduce the risk of harm more than similar features that have similar goals and effects but require significant additional product changes but. **In particular, measure US2 should apply to a more limited set of functionalities (e.g. not to any service that allows posting of content) and only to services that are at high (rather than medium or high) risk of the relevant harms.** In general, the Codes should outline the aim (e.g. empower a user to prevent someone from directly harassing another user) and provide illustrative examples (not explicit

requirements) for how this can be done, rather than mandating blocking and muting. We would ask that **enough flexibility be drafted into the Codes that allow for different measures as long as they are adequately addressing the risk presented**.

Carve-outs for private communications

As also reflected in our response to the Illegal Harms Consultation, in order to adequately protect users' privacy rights and avoid imposing disproportionate burdens on services, we recommend that Ofcom **clearly distinguish between the obligations applicable to private and public communications**. The current draft guidance is at odds with the intention of the Act (see, for example, paragraphs 12(2)(c) and (d) and 13(4) of Schedule 4), and is inconsistent with the approach taken in relation to illegal harms.

A detailed response to your questionnaire follows.

Your response

Question	Your response
Volume 2: Identifying the services children are using Children's Access Assessments (Section 4).	
Do you agree with our proposals in relation to children's access assessments, in particular the aspects below. Please provide evidence to support your view. 1. Our proposal that service providers should only conclude that children are not normally able to access a service where they are using highly effective age assurance? 2. Our proposed approach to the child user condition, including our proposed interpretation of "significant number of users who are children" and the factors that service providers consider in assessing whether the child user condition is met?	Confidential? – N We welcome Ofcom's approach of providing detailed guidance to services on how to carry out a Children's Access Assessment. We have some suggestions on how the guidance could, in our view, be made clearer and more proportionate, which we hope will assist Ofcom. **Approach to the child user condition - "Significant number"** The guidance sets out Ofcom's view that what constitutes a "significant number" of children is context specific and that "even a relatively small absolute number or proportion of children could be significant in terms of the risk of harm to children". This leads to the assumption that "this term should be understood as indicating that the number of children on the service is material in the context of the service in question (i.e. not insignificant in that context)" (Volume 2, para 4.23). We understand that Ofcom's intention is to acknowledge the varying relative sizes and reaches of in-scope services by including an element of proportionality and context-sensitivity into the

Question	Your response
3. Our proposed approach to the process for children's access assessments?	definition of "significant". Ofcom's guidance (Vol 2, para 4.44) suggests that "significant" should be interpreted to mean "most Part 3 services (other than those which can be certain that there are no child users)". However, this does not reflect the statutory definition in s.35(4)(a) of the Act, which states that "significant" means "significant in proportion to the total number of United Kingdom users of a service...". The guidance is therefore incompatible with the legislative intent of the Act and would be a disproportionate burden to many services. The interpretation by Ofcom that "a significant number of children" in s. 35(3)a) of the Act should be understood as meaning that "most Part 3 services that are not using highly effective age assurance" would be captured (para 1.4 of the draft Child Access Assessments guidance) goes beyond the legislative intent of the Act and would also be a disproportionate burden to many services. We consider that there are many other ways in which services that do not deploy highly effective age assurance could still conclude that they are not likely to be accessed by children. Indeed, at other parts of the guidance (e.g. Table 7), Ofcom sets out a range of other factors that service providers can consider in assessing whether their service satisfies the child user condition. Suggested amendment We recommend that Ofcom remove paragraph 1.4 and references to a "relatively small absolute number or proportion" constituting "significant" from the guidance, and instead explicitly recognise that services should make their own determinations on what is significant based on relevant considerations, including the number of impacted child users. We also recommend that Ofcom aligns with the ICO's approach in its guidance on the Children's Code that what is "significant" "depends on a variety of factors relating to the type of service, how it has been designed and the [personal data processing / privacy] risks that it presents to children.". **Where children's access is accepted** The guidance recognises that where a service "conclude[s] that the child user condition is met, you do not need to gather additional evidence or keep a detailed record of the evidence you have relied on to support this conclusion. You must record your outcome" (Annex 5, para 4.5). The guidance also appropriately makes clear that services may, but are not required, to use the template at Annex 5, para A1.8. We support this approach and

Question	Your response
	encourage Ofcom to maintain it. **Highly effective age assurance** We set out our comments on what should be regarded as “highly effective age assurance” in response to Questions 31 to 35 (Section 15) below.
Volume 3: The causes and impacts of online harm to children **Draft Children’s Register of Risk (Section 7)**	

Proposed approach: 4. Do you have any views on Ofcom's assessment of the causes and impacts of online harms? Please provide evidence to support your answer. a. Do you think we have missed anything important in our analysis? 5. Do you have any views about our interpretation of the links between risk factors and different kinds of content harmful to children? Please provide evidence to support your answer. 6. Do you have any views on the age groups we recommended for assessing risk by age? Please provide evidence to support your answer. 7. Do you have any views on our interpretation of non-designated content or our approach to identifying non-designated content? Please provide evidence to support your answer. **Evidence gathering for future work:** 8. Do you have any evidence relating to kinds of content that increase the risk of harm from Primary Priority, Priority or Non-designated Content, when viewed in combination (to be considered as part of cumulative harm)? 9. Have you identified risks to children from GenAI content or applications on U2U or Search services? a) Please Provide any information about any risks identified 10. Do you have any specific evidence relevant to our assessment of body image content and depressive content	Confidential? – N We have included our response to the majority of these questions in our comments on Section 12 below. **Definition of Non-Designated Content (NDC)** Currently, Ofcom has made a preliminary assessment that proposes various categories of content are considered NDC. We note that the Act (s.11(5)-(6); s.13(2); s.28(5); s.30(2)) states that services will identify NDC themselves through the children's risk assessment, rather than Ofcom specifying specific types of content as NDC in its guidance. The approach that Ofcom appears to be currently taking is therefore at odds with the statute, including the appropriate statutory process for adding additional categories of content to PPC and PC, as set out in s.63 of the Act. This is because this approach, in effect, "upgrades" certain kinds of NDC to priority content. It should therefore be Parliament, rather than Ofcom, that designates new categories of priority content in this fashion through the statutory process. We therefore recommend that Ofcom avoids providing specific guidance as to the categories of NDC, and instead aligns its guidance more closely to the process envisaged in the Act, allowing services to make context-specific judgements about relevant content based on an assessment of risk. For the avoidance of doubt, and as discussed further below, we consider that the two categories that Ofcom have currently identified are overly nebulous and imprecise to be incorporated into the requirements, whether by Ofcom or through the statutory process. Furthermore, Ofcom states that it does not currently have sufficient evidence to determine "*the relationship between specific kinds of content, and the material risk of significant harm, and to define the categories of content more clearly*" (Volume 3, para 7.9). Given the significant impact that Ofcom's proposals will have on in-scope services, it is absolutely imperative that Ofcom's recommendations are grounded in thorough evidence, and that there is a clear link between the specified category and significant harm. We therefore support Ofcom's initial conclusion that it requires additional evidence before confirming these categories of NDC. We also consider that it would be preferable to allow some time for the online safety regime to "bed-in" and for services to establish their responses to the existing requirements, before specific categories of NDC are identified and added to existing measures. This will enable Ofcom and services to work together

Question	Your response
as kinds of non-designated content? Specifically, we are interested in: a) (i) specific examples of body image or depressive content linked to significant harms to children, b. (ii) evidence distinguishing body image or depressive content from existing categories of priority or primary priority content. 11. Do you propose any other category of content that could meet the definition of NDC under the Act at this stage? Please provide evidence to support your answer.	to identify further potential categories of NDC, and how best to protect children from it, with the benefit of having evidence available about the implementation of requirements in relation to primary priority content and priority content. We therefore recommend that Ofcom revisit the conclusion drawn from its preliminary assessment that the definition of NDC could encompass "body image content" and "depressive content". These categories are overly broad and subjective. For example, content on the Holocaust, knife crime, or even content on climate change, could be considered to be depressive content, but could also be vitally important for educational purposes. Similarly, whether a video about a workout regime constitutes body image content is heavily context-dependent. The lack of precision in defining the categories means there are significant technical challenges in implementing and enforcing compliance with requirements that relate to NDC. Requiring services to take steps in relation to such subjective and amorphous categories will likely result in a disproportionate regulatory burden on services, and a risk of misalignment with the policy intent. It may also lead to children being inadvertently prevented from seeing non-harmful content that has been misidentified as NDC, limiting important access to information, the ability to build community, freedom of expression, and creativity. If these two categories are to be considered as NDC, further consultation and greater consideration of ways in which these categories of content can be more precisely defined is therefore necessary. We also agree with Ofcom that more evidence is required on whether these content categories meet the "material risk" requirement in the Act. We consider Ofcom should publish this evidence and consult on it before undertaking further policy development in this space. We would be happy to work with Ofcom and the broader industry to find a way forward that is both practical and meets the policy expectations in the Act, to ensure that compliance measures meet the policy intent and balance risk with the protection of children.
Draft Guidance on Content Harmful to Children (Section 8)	

Question	Your response
12. Do you agree with our proposed approach, including the level of specificity of examples given and the proposal to include contextual information for services to consider? 13. Do you have further evidence that can support the guidance provided on different kinds of content harmful to children? 14. For each of the harms discussed, are there additional categories of content that Ofcom a) should consider to be harmful or b) consider not to be harmful or c) where our current proposals should be reconsidered?	Confidential? – N We set out in response to Section 7 our comments in relation to the proposed definition of NDC. We urge Ofcom to be mindful that, due to their scale, many services will be relying on automated systems to make judgments about whether content on the service falls within scope of the categories of content that is harmful to children. As noted below in Section 20, in relation to recommender systems, there are significant technical challenges with making these judgements, especially for certain more subjective content categories. For example, it may be much more challenging to identify "content depicting challenges" at scale, given it is so context-dependent, as opposed to pornography. As Ofcom is aware, there is also a risk of an adverse impact on freedom of expression from over-removal, where content categories are defined broadly and services are expected to err on the side of removal if there is doubt about whether a specific piece of content falls within one of the categories. For example, we note that at paragraphs 8.30-8.32 of Volume 3, Ofcom states that: "We have included considerations about "recovery content" in our suicide, self-harm and eating disorder sections because it may not always be clear whether recovery content does, or does not, meet the definition of PPC. Although some of this content may serve a valid purpose to individuals in allowing self-expression and aiding their recovery, it may still be harmful to children. There is a significant amount of content online that meets the definition in the Act of "content which encourages, promotes or provides instructions for suicide" or "an eating disorder" that in some way is intended to be, or is presented or described as, recovery content." This suggests that Ofcom may be expecting services to err on the side of removal in relation to "recovery content", which risks severely negatively impacting children by restricting their ability to seek support. Although this may be consistent with Ofcom's intention, we would welcome assurance from Ofcom that services have the discretion to consider whether content such as "recovery content" is more helpful or harmful, without being required to pre-judge that it is likely to be harmful.

Question	Your response
Volume 4: How should services assess the risk of online harms? **Governance and Accountability (Section 11)**	
15. Do you agree with the proposed governance measures to be included in the Children's Safety Codes? a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence. b) If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 16. Do you agree with our assumption that the proposed governance measures for Children's Safety Codes could be implemented through the same process as the equivalent draft Illegal Content Codes?	Confidential? – N In relation to the proposed governance and accountability measures, we refer you to our responses to Questions 3 (8.1) to 6 (8.4) of the Illegal Harms Consultation (which proposed near identical measures in relation to the illegal content safety duties). In particular, our suggestions in relation to the proposed measure requiring services to name a person accountable to the most senior governance body for compliance are equally applicable to the equivalent measure in the Child Safety Codes.
Children's Risk Assessment Guidance and Children's Risk Profiles' (Section 12)	
17. What do you think about our proposals in relation to the Children's Risk Assessment Guidance? a) Please provide underlying arguments and evidence of efficacy or risks that support your view. 18. What do you think about our proposals in relation to the Children's Risk Profiles for Content Harmful to Children? a) Please provide underlying arguments and evidence of efficacy or risks that support your view.	Confidential? – Y (partly) **Overarching Comments** We recognise the importance of clear and specific guidance on how to undertake risk assessments in a manner that is consistent with the requirements of the OSA. Google already has well established systems for assessing risk to users across our products and services, largely through a cyclical process of: (i) identifying emerging harms and gaps in existing policies; (ii) gathering examples of how a particular harm has manifested on a service; (iii) developing or updating policies and enforcement guidelines; and (iv) implementing mitigations such as awareness raising measures, product design or process changes; and (v) assessing the impact of the policy or design change, and whether it has addressed the relevant harm. Ofcom's Guidance is therefore likely

Question	Your response
Specifically, we welcome evidence from regulated services on the following: 19. Do you think the four-step risk assessment process and the Children's Risk Profiles are useful models to help services understand the risks that their services pose to children and comply with their child risk assessment obligations under the Act? 20. Are there any specific aspects of the children's risk assessment duties that you consider need additional guidance beyond what we have proposed in our draft? 21. Are the Children's Risk Profiles sufficiently clear and do you think the information provided on risk factors will help you understand the risks on your service? a) If you have comments or input related to the links between different kinds of content harmful to children and risk factors, please refer to Volume 3: Causes and Impacts of Harms to Children Online which includes the draft Children's Register of Risks.	to be particularly important for small businesses that may be less familiar with risk management concepts and may not have existing organisational risk management frameworks. While the processes described in the draft Children's Risk Assessment Guidance at Annex 6 may be helpful for some businesses, in general the approach is overly prescriptive. For mature organisations with existing organisational risk management frameworks, such as Google, it is disadvantageous to be required to align with prescriptive requirements in order to benefit from the baseline "stronger position" as identified in Volume 4, para 12.4. In addition to imposing significant compliance costs, the Guidance is, in places, disproportionate and lacks flexibility (as explained further below). While a more prescriptive approach may be helpful for less mature businesses, for those with established risk management practices there is a risk of undermining the effectiveness of the current risk management processes. Below, we have identified specific aspects of the risk assessment proposals where we believe improvements could be made to the Codes in order to better achieve the policy intent. **Proportionality** The Guidance recognises the relevance of proportionality to the risk assessment process: "How you carry out the children's risk assessment will depend on the size and nature of your service, but it must meet all of the elements specified in the children's risk assessment duties (section 11(6) for U2U services and 28(5) for search services)" (Annex 6, para 2.7). However, the principle is not consistently applied in the Guidance. In particular Ofcom's approach to evidence is disproportionate. This includes the assumption that the data described as "core evidence" data will be readily available to every service (Annex 6, para 2.11(a)). Some data, such as quantitative data surrounding user complaints and content moderation, are not necessarily currently collated by services across the industry in a manner that allows for assessment as to specific harms or offences in a reliable manner. Most data that is collected may not be readily usable to derive reliable UK-specific insights. Our view is that in some cases it is not proportionate to the risk to collect that data in a detailed

Question	Your response
	and granular form for all services. In other cases, it may take some time to gather the required data. The form and extent of the data to be collected should be proportionate to the risk profile of the service, and services should be deemed to have completed a "suitable and sufficient" risk assessment, while using reasonable substitutes or proxies for "core evidence", during an initial implementation period (for example, the first year during which services must complete a child risk assessment). <u>Suggested amendment</u> We welcome that Ofcom has taken a more nuanced and proportionate approach to the responsibilities of service providers in other areas. For example, Ofcom's draft Codes and Guidance apply different measures for different services depending on whether the service is a "large service" or a "multi-risk" service (among other factors). We also welcome that Ofcom has recognised the ability of services to take a proportionate approach to determining when and if it is necessary to rely on enhanced evidence. However, Ofcom's guidance should also recognise that services should take a proportionate approach to determining the appropriate nature, extent, and form of both core and enhanced inputs. We recommend that Ofcom revise the Risk Assessment Guidance in a similar way with regards to "core evidence" and "enhanced evidence" to allow a proportionate and multi-tiered approach to what evidence is relevant for a particular service. In addition, we would ask that Ofcom gives consideration to a grace period for some of the core evidence where it will take services time to build up UK specific data. **"Significant change"** We recognise the importance of keeping risk assessments up to date and assessing how significant changes may impact risk. However, Ofcom's current Guidance concerning when a change will amount to a "significant change" covers an overly broad range of circumstances and risks going well beyond the scope of the OSA. We refer to and re-iterate the comments and proposed amendments that we made in response to the Illegal Harms Consultation on this topic in response to Question 7. While parts of the Child Safety Consultation (e.g. at Vol 4, para 12.99) recognise that it would not be proportionate to capture routine

Question	Your response
	updates and system changes, this principle is not reflected in the operational parts of the Guidance. **Prejudgment of outcome** The Guidance contains a number of statements that appear to prejudge the outcome of particular courses of action or assume points that may or may not be true. <u>Child risk assessment</u> In relation to risk assessments in general, Ofcom states that "conducting a suitable and sufficient children's risk assessment should result in a reduction of the risk of harm to children on the service" (Volume 4, para 12.28), and similarly that "the findings of each risk assessment will each lead service providers to put in place different Codes measures to reduce the risk of each kind of content harmful to children or illegal content" (Annex 6, para 4.18). These statements appear to assume that a service provider is not already implementing processes which will be sufficient to reduce the risk of harm. This conflicts with Ofcom's recognition that a service provider may already have "some measures in place to reduce the risk of harm to children" (Annex 6, para 4.74). While we understand that this assumption may be true in some cases, Google already has well-established systems for assessing and mitigating risk to users across our products and services. It is therefore possible that the measures in place mean that the child risk assessment process does not identify further measures that are required to reduce the risk of harm to children on a service. We would urge Ofcom not to prejudge the outcome of the assessment. <u>Enhanced evidence</u> The Guidance appears to assume that it will be necessary and appropriate for larger services to have regard to enhanced inputs without reference to the adequacy of core inputs, the nature of the service, or the specific risk being assessed (e.g., at Annex 6, para 4.36). It is disproportionate to assume that, for larger services, it will not be possible to carry out a suitable and sufficient risk assessment without recourse to enhanced evidence, especially with respect to all potential harms to be assessed. <u>Suggested amendments</u> ● Ofcom should ensure that the outcome of a course of action is not assumed. Ofcom could recognise that

Question	Your response
	services may have already successfully implemented mitigation measures, and therefore the risk assessment may not automatically bring further reduction of harm. ● Remove the assumption that larger services are required to use enhanced inputs irrespective of the sufficiency of other evidence, the nature of their service, or risk profile. **Child-specific evidence and the impact on privacy** Safety is core to how we develop and operate our services, and we understand our responsibility to keep users safe while protecting their privacy and promoting the free flow of information. However, we have some concerns that the requirements in the current draft Guidance have a disproportionate impact on children's privacy, in particular regarding the potential requirement to collect and process additional personal data in connection with children's risk assessments. In response to the Illegal Harms Consultation on this topic, in response to Question 7, we set out our concerns about Ofcom's approach to the collection and use of user data, which did not recognise that it would be rare for a service to have the level of information Ofcom assumes to exist about their user base demographics or a particular user's characteristics. We also explained the need for providers to consider carefully whether the processing of data is compliant with that service's obligations under the UK GDPR and Data Protection Act 2018. Our primary concern is that the Guidance continues to assume that services will have more detailed information about their user base than they do. The Guidance goes beyond the statutory requirements and recommends the collection and use of data about children in a way that raises the following potential privacy concerns and operational challenges: 1. We take seriously our privacy obligations, including the data minimisation principle in the UK GDPR and its specific application to children in the ICO Children's Code, which Ofcom references (Annex 6, para 4.44). This principle requires services to collect and retain only the minimum amount of personal data that is required to provide the elements of the service with which a child is actively and

Question	Your response
	knowingly engaged. We would expect that the Guidance for services to use the "best available evidence" does not create a new expectation for services to collect additional personal data beyond what is already being collected by services. 2. Operationally, Ofcom's requirement to consider five specific age "buckets", rather than the effect of harm to children in different age groups (as identified by the service in its most recent risk assessment), as set out in the OSA, will require the collection and analysis of evidence and data in those specified buckets, which is disproportionate because it requires services to implement additional systems and processes without a clear commensurate reduction in risk. It is also not clear that new collection and processing of personal data would be necessary and proportionate to the related benefits of online safety, as required by the UK GDPR and, in particular, whether the level of differentiation proposed will be necessary and proportionate for all services and in respect of all risks. Children of the same age, or even of different ages within the age bands proposed by Ofcom, may have very different levels of maturity, ability, and awareness, and interpersonal or cultural dynamics may vary considerably from one family to the other. A measure that works for one particular child user may not work for another. For that reason, we recommend that Ofcom implement an approach that gives services latitude to consider a wide spectrum of potential risks and impacts, rather than a one-size-fits-all approach. The current proposals appear to suggest a limited set of measures that are suitable for one type of user, that are to be applied wholesale to all other child users. <u>Suggested amendments</u> ● Make clearer that the "best available" standard must be interpreted having regard to obligations under UK GDPR and the ICO Children's Code, and does not create an expectation that services collect additional personal data beyond what they are already collecting. ● Ofcom should consult with the Information Commissioner's Office to revise the Guidance to provide more clarity to providers around the permissibility of collecting and processing children's data for the purposes of carrying out risk assessments.

Question	Your response
	What evidence is sufficient to assess the number of children users for any given service should be determined having regard to the nature of the service and the principle of proportionality. While the Guidance might appropriately highlight potential limitations that may apply to certain methodologies for calculating the number of children users, the current Guidance risks prejudging and/or artificially augmenting risk assessments without justification. <u>Suggested amendments</u> ● Remove reference to the "strong evidence to demonstrate otherwise" standard. ● Remove the instruction to "assume that the true number of children using their service… could be much greater". **Harm Threshold Assumptions** The Risk Level Table (Annex 6, table 4.4) identifies the mere number of child users as a potential factor contributing to impact. Google agrees with Ofcom's Guidance (Volume 4, para 12.49) that "evidence on number of children – like any other element in the risk level table – is only one of various risk factors that services should consider as they determine their risk level and how best to mitigate it, and that in some instances it may be a weak indicator of risk levels. It is possible for a service with a large number of children to be low risk, and for a service with a small number of children to be high risk, depending on the specific circumstances of the services". While we agree that the number of child users may contribute to the scale of a risk, there is no basis in any of the OSA, principle or best practice for Ofcom's suggestion (at Volume 4, para 12.50) that "...any service with more than 1 million (or between 100,000 and 1 million) monthly UK child users would need a range of robust evidence to demonstrate that it does not in fact pose high (or medium) risk of harm to children in respect of a given kind of content." There is a similar statement made in the footnotes to Annex 6, table 4.4. While Ofcom explains (at Volume 4, para 12.44 to 12.46) that the numerical range represents 0.7 to 7% of the UK child population, it is not clear why this logic applies to risk assessments. For example, this statement is inconsistent with Ofcom's acknowledgement that in "some instances, the number of children may be a weak indicator of risk level, and service providers should consider this indicator alongside other factors" (Annex 6, table 4.4). While such a service would need sufficient

Question	Your response
	evidence to ensure its risk assessment was suitable and sufficient, the attempt to impose a different evidential standard is confusing and appears to be detached from the intended purpose. Suggested amendment ● Delete Volume 4, para 12.50 and amend Annex 6, table 4.4 to remove the presumption that any service with more than 1 million (or between 100,000 and 1 million) monthly UK child users would pose high (or medium) risk of harm to children in respect of a given kind of content. **Approach to level of harm** Ofcom has presented various approaches to measuring harm, including consideration of reach of the service, cumulative harm, indirect harm. Cumulative harm The Guidance notes that providers should consider how features or functionalities impact the risk of cumulative harm for the risk assessment, including a consideration of how any particular features or functionalities affect how much children use the service. Cumulative harm is defined as when: 1. content is repeatedly encountered by an individual; or 2. a particular kind of content is encountered by an individual in combination with a different kind of content (Annex 6, para 4.27). Ofcom provides the rationale that where children are spending more time on services, it could increase the risk of cumulative harm (Annex 6, paras 4.29 - 4.30). We consider that, in relation to the first limb of the definition in particular, Ofcom should clarify that services should take a proportionate approach when assessing what constitutes cumulative harm, such that the harm that results from the content being repeatedly encountered is proportional to the number of times the content encountered. We recommend that Ofcom explain that an effective approach to cumulative harm will also likely involve the assessment of both remediability and probability of a given risk. Remediability allows a provider to consider the gravity of a harm (e.g. the risk associated with impacts to the physical, mental, or material well-being of an individual) and whether the individual can be returned to their original state

Question	Your response
	before the harm. The risk that an individual may repeatedly encounter a specific type of content is also considered when determining remediability. When assessing probability of the impact occurring, a service considers frequency of the risk occurring (e.g., prevalence), which indicates the likelihood that content is repeatedly encountered by an individual. In both cases, assessment of cumulative harm is difficult to measure quantitatively and may require qualitative insights from such groups as external stakeholders and/or subject matter experts. *Indirect harms* According to the Guidance, indirect harm occurs when children are harmed, or the likelihood of harm is increased, as a result of another individual who views the content by them doing or saying something to that other child as a result of viewing the content (Annex 6, para 2.4). We appreciate that there may be several ways in which a service could go about seeking to measure indirect harm, recognising the challenges in seeking to do so. [✂] Ofcom does not provide guidance on how it considers services should consider indirect harm or the circumstances in which they should do so. We appreciate that there may be several ways in which indirect harm could be measured: it would be helpful if Ofcom could further clarify that harm may be sufficiently measured without taking into account indirect harm as a distinct category, and that services should use their discretion to determine whether harm has been sufficiently assessed, including whether and how a suitable and sufficient risk assessment with respect to any given category of risk requires an assessment of indirect harm.
Volume 5 – What should services do to mitigate the risk of online harms **Our proposals for the Children's Safety Codes (Section 13)**	
Proposed measures 22. Do you agree with our proposed package of measures for the first Children's Safety Codes? a) If not, please explain why.	Confidential? – N We are broadly aligned with the approach that Ofcom have taken in the draft Codes and appreciate the difficulty of creating rules that can be applied effectively across the range of services who may fall into scope. However, one issue with the current draft Codes, which threatens to undermine the security and privacy of

Question	Your response
Evidence gathering for future work. 23. Do you currently employ measures or have additional evidence in the areas we have set out for future consideration? a) If so, please provide evidence of the impact, effectiveness and cost of such measures, including any results from trialling or testing of measures. 24. Are there other areas in which we should consider potential future measures for the Children's Safety Codes? a) If so, please explain why and provide supporting evidence.	individuals, is a failure to distinguish between public and private communications. Unlike the approach taken in the draft Illegal Harms Codes, the current Guidance appears to require that measures are applied equally to private and public communications. The current draft Guidance is at odds with the requirements of Schedule 4 of the Act, which mandate that when drafting the Codes of Practice, Ofcom must have regard to users' rights of freedom of expression and protecting the privacy of users (see Schedule 4, para 10(2)), and must consider the levels of risk to children on a service, and the potential impact on freedom of expression, before recommending a requirement to use age assurance (Schedule 4, paragraph 12(2)(c) and (d)), and to only recommend a proactive technology measure in relation to public, not private, communications (Schedule 4, paragraph 13(4)). We therefore recommend that Ofcom align the draft guidance with the approach taken in the Illegal Harms Codes, by distinguishing between public and private communications, due to the statutory obligation to balance users' rights of privacy with the use of safety measures.

<table>
<tr><td colspan="2">Developing the Children’s Safety Codes: Our framework (Section 14)</td></tr>
<tr><td>25. Do you agree with our approach to developing the proposed measures for the

Children’s Safety Codes?

a) If not, please explain why.

26. Do you agree with our approach and proposed changes to the draft Illegal Content Codes to further protect children and accommodate for potential synergies in how systems and processes manage both content harmful to children and illegal content?

a) Please explain your views.

27. Do you agree that most measures should apply to services that are either large services or smaller services that present a medium or high level of risk to children?

28. Do you agree with our definition of ‘large’ and with how we apply this in our recommendations?

29. Do you agree with our definition of ‘multi-risk’ and with how we apply this in our recommendations?

30. Do you agree with the proposed measures that we recommend for all services, even those that are small and low-risk?</td>
<td>Confidential? – N

In relation to the definition of a ‘large’ service, we reiterate the points made in our answer to question 14 of the Illegal Harms Consultation. We consider that the bar for large services is set too low and would flag the risk of overcounting users. We would welcome both a higher threshold and Ofcom expressly providing flexibility for services about how users are counted, such that providers have discretion to determine who is a “user” by reference to whether there is a realistic prospect of the person being exposed to harm.

We would also welcome a shorter period for assessing monthly users (e.g. 6 months rather than 12) and an ability for services to ‘self-certify’ that they meet the threshold for a large service, without having to provide precise user counts.

Similarly, in relation to the definition of ‘multi-risk’, we wish to reiterate the points made in our answer to question 15 of the Illegal Harms Consultation. The proposal that any service that is at least medium risk in relation to at least 2 kinds of harmful content is multi-risk and thereby subject to additional obligations is disproportionate, particularly given the very broad definition of non-designated content.

Multi-risk should be limited to services that are high risk in relation to at least 2 types of primary priority and/or priority content. Without this change, there would be insufficient delineation between the treatment of services that are medium risk for content harmful to children and services that are high risk for that harm.</td></tr>
<tr><td colspan="2">Age assurance measures (Section 15)</td></tr>
</table>

31. Do you agree with our proposal to recommend the use of highly effective age assurance to support Measures AA1-6? Please provide any information or evidence to support your views. a) Are there any cases in which HEAA may not be appropriate and proportionate? b) In this case, are there alternative approaches to age assurance which would be better suited? 32. Do you agree with the scope of the services captured by AA1-6? 33. Do you have any information or evidence on different ways that services could use highly effective age assurance to meet the outcome that children are prevented from encountering identified PPC, or protected from encountering identified PC under Measures AA3 and AA4, respectively? 34. Do you have any comments on our assessment of the implications of the proposed Measures AA1-6 on children, adults or services? a) Please provide any supporting information or evidence in support of your views. 35. Do you have any information or evidence on other ways that services could consider different age groups when using age assurance to protect children in age groups judged to be at risk of harm from encountering PC?	Confidential? – N Having a better understanding of the age of our users is only one of the ways in which Google works to ensure that children and teens have safer and more appropriate experiences on our products. Verifying users' ages is not a silver bullet and will not in and of itself lead to age-appropriate experiences for children and teens; it is one of a number of tools we can deploy to keep children and teens safe, and respect their rights to privacy and access to services and information. The current draft guidance on HEAA is overly focused on data-intrusive forms of age verification that rely on "hard identifiers" (as discussed in the ICO Children's Code), without sufficient consideration of the risks that this approach may bring. As noted in the Summary above, this is at odds with the requirement in Schedule 4, paragraph 12(d) of the Act that the Codes of Practice must have regard to "the principle that more effective kinds of age assurance should be used to deal with higher levels of risk of harm to children". As discussed further below, Ofcom's approach should be brought in line with data minimisation best practices, and greater consideration given to children's fundamental privacy rights. The current approach taken in the draft guidance risks alienating vulnerable users by requiring all services to implement intrusive methods of age assurance. The method for establishing the age of the users, and for protecting those users, needs to be proportionate to the risk at hand. We work to reduce the risk for all our users, thus reducing the need to apply more data intensive methods to establish the age of the user. This helps provide a balance between user privacy, data security, and inclusiveness/fairness. We think that age assurance should be used alongside a variety of other tools that we use to keep children safer, such as: a. parental controls such as Google's Family Link that enable parents to better control their child's online experiences b. our mechanisms to detect and remove violative harmful content; c. content filtering tools (such as SafeSearch filter and SafeSearch blur) d. our use of curated spaces for children (such as YouTube kids);

	e. our use of ranking and recommender systems to help connect children to high quality content while minimising the chances they'll see low quality content We think the use of age assurance must be proportionate and risk-based and the measures in the draft Code of practice currently take a blanket approach to age assurance which is not sufficiently calibrated by reference to the level of risk. The approach to age assurance taken by Ofcom must be consistent with the need to preserve users' access to information and services, respect their privacy, and preserve the potential for anonymous or pseudonymous experiences. This may be especially relevant and important for issues such as seeking advice about abusive parents, understanding or exploring sexuality and identity, or seeking confidential help and support. Therefore, the process should involve no more than the minimum level of interference necessary with users' privacy and should therefore avoid the collection or processing of additional personal information, or impinging on the ability of adults to access information. In particular, we are conscious that not all users have access to credit cards or have government issued IDs. Individuals from marginalised or vulnerable groups may be unable or unwilling to share this information, further compounding their disadvantage. We have a number of concerns with the way Measures AA1-6 are currently designed in the draft Code: a. **Meaningful differentiation between risk levels:** The Code requires the use of HEAA where the service is medium **OR** high risk for one or more kinds of Primary Priority Content/Priority Content (depending on the measure). As such, and as also flagged in our response above at Section 14 regarding the definition of 'multi-risk', there's no meaningful differentiation between what type of age assurance is appropriate for services that are 'medium' risk for a particular type of harm and those that are 'high' risk for that harm. b. **Sliding scale of HEAA for different services:** Google prioritises a principled and risk-based approach to age assurance, which we laid out in our Legislative Framework to Protect Children and Teens Online. We acknowledge that certain content and products require

	higher certainty of age. However, it is not clear, either from the draft Code or the draft guidance on HEAA, whether Ofcom interprets the “highly effective” threshold to differ depending on the service type, the nature of the content on the service, and the level of risk presented to children on the service. In other words, what is “highly effective” should be different according to the risk level of a particular service. As pointed out in our Legislative Framework document, there are a range of age assurance methods, which have a differing impact in terms of intruding on user privacy through data collection and use and restricting adult users’ access to important information and services. More data-intrusive methods (such as age verification with “hard identifiers” like photo IDs) should only be expected from services that are high-risk for the most harmful content (i.e. primary priority content such as pornography). c. **Age inference and “challenge age” approach**: We know that the more certainty and granularity required about a user's age, the more intrusive and data intensive the method of age assurance, and the higher the impact on users’ ability to access services and information. This is why we think it is important that the Act allows providers of user-to-user services to use age verification **or** age estimation to comply with the child safety duties. We think it is important that, consistent with the Act, Ofcom allows service providers the flexibility to use age inference models, where they are sufficiently accurate. Where these models are used, it would be appropriate for Ofcom to require the model to operate within a reasonable level of accuracy, as opposed to imposing a specific accuracy range, which may be disproportionate given the risk that the service poses to users. If a service has reasonable grounds to conclude based on age estimation that a user is an adult, services should be able to treat this user as such.
Content moderation U2U (Section 16)	

36. Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views. 37. Do you agree with the proposed addition of Measure 4G to the Illegal Content Codes? a) Please provide any arguments and supporting evidence.	Confidential? – N We reiterate the comments made in our response to the Illegal Harms Consultation, particularly in the answers to Question 18 and 19. As with the draft Illegal Harms Codes, in our view the draft Children's Codes are, in parts, overly prescriptive and do not provide platforms with the necessary flexibility to innovate and implement changes at the speed with which bad actors operate. In particular: a. General monitoring: We would encourage Ofcom to explicitly set out that none of the measures in the draft code would oblige platforms to generally monitor the service for content that is harmful to children - as has been confirmed from the despatch box (see references in the Summary above). b. Performance Targets: Measures CM3 and SM4 require services to set and record performance targets covering at least the time that content harmful to children remains on the service before it is taken down and the accuracy of the decision making. These metrics may be helpful to organisations new to their content moderation journey. However, rather than serving as illustrative examples, they are currently captured as requirements in the code that would require us to maintain separate Trust & Safety metrics for content that is violative of our Community Guidelines, thus hindering existing processes and mechanisms to evaluate the efficacy of existing content moderation measures. We would urge Ofcom to provide flexibility to allow the varying different services to which the Codes apply to use a range of performance metrics that most effectively address the issue outlined. c. Prioritisation framework: Measures CM4 and SM5 require large or multi risk services to use a prioritisation framework that includes Ofcom-specified factors, like an assessment of the virality of the content and the severity of the content. Google currently uses prioritisation frameworks, which include a number of factors, but the policies adopted depend on the type and severity of harm and the type of product, to reflect the differing functionality and risk profiles. We would suggest Ofcom provides more flexibility by requiring services to have appropriate prioritisation frameworks,

	providing examples of metrics rather than prescribing them.
Search moderation (Section 17)	
38. Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views. 39. Are there additional steps that services take to protect children from the harms set out in the Act? a) If so, how effective are they? 40. Regarding Measure SM2, do you agree that it is proportionate to preclude users believed to be a child from turning the safe search settings off? The use of Generative AI (GenAI), see Introduction to Volume 5, to facilitate search is an emerging development, which may include where search services have integrated GenAI into their functionalities, as well as where standalone GenAI services perform search functions. There is currently limited evidence on how the use of GenAI in search services may affect the implementation of the safety measures as set out in this code. We welcome further evidence from stakeholders on the following questions and please provider arguments and evidence to support your views: 41. Do you consider that it is technically feasible to apply the proposed code measures in respect of GenAI functionalities which are likely to perform or be integrated into search functions?	Confidential? – N We note that, under the proposed measures, search services must downrank and/or blur primary priority content unless the service has reasonable grounds to believe the user is an adult. Services should also consider if it is appropriate to downrank and/or blur Priority Content. Due to the difficulty of estimating the age of signed-out users, we are concerned that these measures will be difficult to implement in practice, without requiring us to take a blanket approach of blurring or downranking the content for all signed out users. This would amount to a significant interference with adult users' rights to access information. With regard to SM2, our current SafeSearch functionality allows supervised accounts - such as those operated by a school administrator or by a parent under Family Link, to lock the SafeSearch setting on. This gives those managing the supervised account flexibility to determine what settings are appropriate for their users and/or family members. Per Google policies, users under the age of consent must use a supervised account. Accounts not under supervision are given the option to turn the setting off; however we default SafeSearch filter on for signed-in accounts when our systems indicate that the users may be under 18, consistently with the proposed measure. Additionally, our SafeSearch blur setting is defaulted on for all users who do not already have the SafeSearch filter enabled. Given that many signed out users are over the age of 18, we believe that the current functionality of our SafeSearch features provide appropriate balance between ensuring user safety and access to information and recommend that signed out users continue to have the flexibility to turn the SafeSearch filter off if desired. SM4 requires service providers to set performance targets for, and assess, "the time that search content that is content that is harmful to children remains on the service before it is actioned".

42. What additional search moderation measures might be applicable where GenAI performs or is integrated into search functions?	We do not think this measure is technically feasible given this would require general monitoring of the service. If this is a reference to the time the content remains on the service after Google becomes aware of it, we consider that this is not a suitable performance target measurement for search engines for harmful content. This is because the length of time that potentially harmful content remains on the service before a complaint is processed depends on a large set of circumstances and differs greatly depending on the type of content. Setting a target time is therefore not particularly effective. For Google Search, a performance target related to the time taken to process complaints becomes particularly challenging when it applies to downranking, as it is difficult to specify to what extent a result is downranked with precise attribution as to why that downranking occurred, as it could occur for non-associated reasons, including changes to our ranking algorithm. As explained in our response to the Illegal Harms Consultation (see our answer to Question 19), it is not clear to us what “downrank” means under the draft Code, especially given that the same page might rank differently for a variety of different queries. For example, if a user searches for “UK regulator”, Ofcom's website appears as the fifth result. However, if you search for “UK digital safety regulator”, it appears as the first result. Does the first scenario qualify as downranking only because the first four results had a higher quality score, potentially because they provide a list of different UK regulators and, therefore, might be considered more relevant? We recommend clarification of the definition of “downranking” to require ranking algorithms to be altered only in circumstances where service providers think it is possible or appropriate; or, in the alternative, in cases of general queries, rather than navigational queries targeted to finding a particular page or site. Moreover, when speaking about pages with potentially harmful but legal content, applying a penalty to a page might result in it not appearing highly in search results for a general query; but the search content might rank more highly in response to “navigational” queries, that are targeted to finding a particular page or site. It’s therefore not clear what “downranking” means in a context where the query has one obviously correct answer. We also think providers should be given the choice as to

	whether to blur or downrank without having to assess the listed factors (for example, they should be able to blur in all cases, where appropriate). We would suggest that the factors that service providers should have regard to when deciding whether to blur or downrank should not be prescriptively set out. Examples of factors that may be considered could include the prevalence of content that is harmful to children hosted by the website and the interests of users, but this should not be mandated. PCS B1.5 could therefore state that: ● "The provider may (where relevant based on the nature of the content) have regard to the factors set out in PCS B1.5:"; or ● Ofcom could replace (a) - (c) of PCS B.15 and instead say that "In considering whether to blur or downrank the search content concerned, the provider should have regard to appropriate factors which assess the harmfulness of the relevant search content and the volume and nature of non-harmful material that would be affected".

User reporting and complaints (Section 18)	
43. Do you agree with the proposed user reporting measures to be included in the draft Children's Safety Codes? a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence. b) If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 44. Do you agree with our proposals to apply each of Measures UR2 (e) and UR3 (b) to all services likely to be accessed by children for all types of complaints?	Confidential? – N We refer to our response to Question 28 (16.1) of Ofcom's Illegal Harms Consultation in respect of this question, in particular in relation to the requirements to provide an indicative timeframe for a response in acknowledgements of complaints (UR3); and our position in relation to spam and providers of malware (UR4). In addition to this, we note that Measure PCU C6 (UR4(b)(i)) sets out a number of matters to which service providers should have regard when determining what priority to give to review of a relevant complaint which is an appeal. This wording is reflected in the consultation on amendments to the Illegal Content Codes of Practice published on 8 May 2024, and our response applies to both proposals. We consider it is appropriate to require a prioritisation framework, but there should not be any requirement to include specific factors in the framework, as these may not be applicable to all harms and all products. We therefore

a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence. b) If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 45. Do you agree with the inclusion of the proposed changes to Measures UR2 and UR3 in the Illegal Content Codes (Measures 5B and 5C)? a) Please provide any arguments and supporting evidence.	recommend that the Codes allow for more flexibility for services to prioritise appeals as is most appropriate for their systems, rather than introducing unnecessarily burdensome requirements that could lead to the opposite effect to that which is intended. Where services already handle appeals on a timely basis (for example, see the median time to action an appeal at Table 4.4.1 of our EU DSA Transparency Report), introducing general obligations applicable to all services and harms may unhelpfully slow down the review system. Ofcom could therefore recommend that where all appeals are already dealt with promptly, Ofcom's suggested prioritisation framework need not apply. **Complaints and Appeals (Search) - Annex 8** Spam complaints and providers of malware We note that Measure UR3 requires service providers to acknowledge receipt of complaints and provide an indicative timeframe for a decision in the acknowledgement. UR4 sets out the appropriate action to be taken in relation to various types of relevant complaints. We do not consider that all complaints should result in an action or an acknowledgment, as some complaints are spam, submitted by bad actors, or lack adequate information to enable action to be taken. We are committed to dealing with all genuine complaints. However, based on our experience across various services, we also foresee a large number of non-genuine "spam" complaints. In our experience, bad actors seek to exploit notice and complaints systems to obtain feedback that will enable them to circumvent detection systems. Recognising this risk, the Digital Services Act appropriately includes an exception for "deceptive high volume commercial content". While at least some spam and complaints by providers of malware will not be a valid complaint within the meaning of the Act, there are other instances where this may be less clear. The draft Codes do not presently recognise that "appropriate action" and acknowledgments should not be necessary in respect of spam, and complaints by providers of malware should be materially different to genuine complaints. Suggested amendments:

	We recommend amending the Codes to clarify that relevant user reporting obligations do not apply to spam or complaints by providers of malware. <u>Reversing decisions on appeal</u> The draft Code states that if a service reverses a decision on appeal, the search content should be restored to its previous position (see PCS C8.2). However, search results are dynamic and the ranking is always changing in response to new web results and information about what results users find helpful. For example, if an appeal decision finds that a content item was not content that is harmful to children, the provider is asked to restore the search content to the position it would have been in had it not been judged to be content that is harmful to children. But it might be the case that underlying factors have changed the impact of the ranking, for example that new web results have been added (or removed) or we know more about what users find helpful. Furthermore, even if the content is not determined to meet the definition of content that is harmful to children, it might be policy violative and the service provider should still retain the ability to downrank content if it is harmful or of low quality, if it violates content policies. In relation to the equivalent requirement for user-to-user services (see PCU C8.2), we note that the same comments apply for services that use recommender systems or other mechanisms that change the order or position in which users encounter content. As with search providers, other underlying factors may have changed where that content would now sit, for example the addition of newer comments on a video, and so restoring the content to the exact position is not possible. Furthermore, the position should be aligned between the Child Safety Codes and the recent amendments made to the Illegal Content Codes. The Code also states that if a decision is reversed on appeal, the relevant moderation guidance should be amended; and any automated technology should be amended where applicable to prevent similar issues. While we welcome the inclusion of the applicability standard in relation to automated content moderation technology, we think this should also be extended to the content moderation guidance. Furthermore, for the avoidance of doubt, the Codes should explicitly state that this should not apply to individual cases but rather should be based

	on an aggregated assessment (e.g. a spike in successful appeal rates). An individual false positive is not necessarily indicative of a systemic issue that requires algorithmic changes. Also, many individualised cases are simply down to fact-specific evaluations of the case at hand, and reversals happen as a result of a better appreciation of the specific content and its context. In those cases, no change to the Guidance is warranted. Suggested amendment: ● An amendment should be made to PCS C8.2 and PCU C8.2, to align with the changes proposed on the consultation on Amendments to Illegal Content Codes of Practice, as published on 8 May 2024, that the provider should "reverse the action taken against the user or in relation to the content (or both) as a result of that decision (so far as appropriate for the purpose of restoring the position to what it would have been had the decision not been made". ● The Code should specify that search services have the discretion not to reinstate content, if it violates content policies or Terms of Service. For example, services should retain the right to demote the content in question if they determine based on the review that the content in question is of low quality, even if it is not harmful to children, or due to changes in ranking algorithms. ● Ofcom should extend the "where applicable" wording in subsection c) to subsection b). Further, for the avoidance of doubt, the Code should clarify that the obligation to adjust content moderation guidance and automated content moderation technology should be based on an aggregated assessment of cases rather than on a case-by-case basis. <u>Complaints system</u> In relation to PCS C2.2 and PCU C2.2 regarding the number of steps/clicks necessary to submit an appeal or complaint, Google does extensive UXR testing to ensure our flows are as user friendly as possible. In our view, the appropriate metric should not be "as few number of clicks as possible" but how intelligible a reporting flow is to users. Focusing solely on the number of clicks creates unintended consequences such as the poor design of the user interface that would in fact discourage reporting or dramatically increase the number of erroneous user reports.

	Suggested amendment: We recommend amending PCS C2.2 to say that the process should be 'as user-friendly as reasonably possible' rather than 'as few steps as reasonably possible' as often having additional steps create clearer and more actionable reports, benefiting both platforms and users. Appeals for downranking (Search) We consider this provision to be problematic. As currently drafted, the draft Code could allow every webmaster whose site is not listed as the first result to have the right to file an appeal. As a practical example, if our search ranking tools downrank sites for legitimate reasons, this mechanism would allow webmasters to abuse OSA appeals as a way to litigate their quality scores. This opportunity for bad actors to game our ranking protections would undermine search ranking quality and safety. For more information please see our response to Q28 of the illegal harms consultation response.

Terms of service and publicly available statements (Section 19)	
46. Do you agree with the proposed Terms of Service / Publicly Available Statements measures to be included in the Children's Safety Codes? a) Please confirm which proposed measures your views relate to and provide any arguments and supporting evidence. b) If you responded to our illegal harms consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 47. Can you identify any further characteristics that may improve the clarity and accessibility of terms and statements for children? 48. Do you agree with the proposed addition of Measure 6AA to the Illegal Content Codes? a) Please provide any arguments and supporting evidence.	Confidential?
Recommender systems (Section 20)	
~~49.~~ Do you agree with the proposed recommender systems measures to be included in the Children's Safety Codes? a) Please confirm which proposed measure your views relate to and provide any arguments and supporting evidence. b) If you responded to our illegal harms consultation and this is relevant to your response here, please signpost	Confidential? – N **Scope** We recommend amending the definition of 'Recommender Systems' to clarify that it only covers promotion of third-party content to users. We understand that it is not the intention of Ofcom or Parliament to restrict how a user's own content is recommended to them within their library / account (e.g. Memories in Photos). For example, in its advice to Government on the categorisation of services, Ofcom noted that the risk posed by content recommender systems was that they "*amplify the breadth, scale and speed of content dissemination on a service by proactively disseminating*

to the relevant parts of your prior response. 50. Are there any intervention points in the design of recommender systems that we have not considered here that could effectively prevent children from being recommended primary priority content and protect children from encountering priority and non-designated content? 51. Is there any evidence that suggests recommender systems are a risk factor associated with bullying? If so, please provide this in response to Measures RS2 and RS3 proposed in this chapter. 52. We plan to include in our RS2 and RS3, that services limit the prominence of content that we are proposing to be classified as non-designated content (NDC), namely depressive content and body image content. This is subject to our consultation on the classification of these content categories as NDC. Do you agree with this proposal? Please provide the underlying arguments and evidence of the relevance of this content to Measures RS2 and RS3. • Please provide the underlying arguments and evidence of the relevance of this content to Measures RS2 and RS3.	*content to **new users or groups of users**"* (paragraph 3.19, emphasis added). **<u>Inclusion of Non-Designated Content</u>** We have concerns about Ofcom's proposal to extend Measures RS2 and RS4 to two potential types of non-designated content: "body image content" and "depressive content", which we set out in response to Section 7 above. More broadly, we are conscious that complying with the recommender system measures will inevitably require using automated technology to detect Primary Priority Content, Priority Content and certain non-designated content at scale. This can be much more challenging for certain more subjective content categories, particularly if they are interpreted broadly. For example, it may be much more challenging to identify "content depicting challenges" at scale, given it is so context-dependent, as opposed to pornography. Furthermore, as explained above, Ofcom fails to distinguish between the age assurance obligations (including in relation to recommender system measures) that are applicable to services that are high and medium risk respectively. This disproportionate approach applies the same burden to services that pose different levels of risk, which results in unduly onerous obligations for medium risk services. It may also have the unintended consequence of reducing the incentive for high risk services to reduce risk. A more proportionate approach would be to interpret "highly effective" in a context-based way, such that different obligations are required of services, depending on their level of risk. This would be consistent with the intention of the Act, which requires services to use proportionate systems and processes to achieve the requirements, and recognises the importance of accounting for different levels of risk of harm to children. This broad-brush approach is also taken in relation to categories of content, as the requirements are triggered for all categories of priority content, even if the service is only at medium risk of one type of priority content. The obligations should, instead, be proportionate and dependent on which type of content is high risk, and whether there are multiple types of high risk content.

	As set out in the Summary above, 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, and is not required by the Act. YouTube RS3 - PCU F3.4 - In relation to YouTube, we recognise the principle that a child user (or users in general) should be able to express negative sentiment towards a *specific piece* of regulated UGC, and that this should result in lower prominence for that content. In general, we use a number of signals which build on each other to help inform our system about what a user finds satisfying. These include, but are not limited to, clicks, watchtime, survey responses, sharing, likes, and dislikes. We shared a blog post with a deeper dive on how recommendations work which you can find here. Under RS3, platforms are also required to give lower prominence to any other piece of content that 'shares significant characteristics' with content that a child user has signalled negative sentiment towards. This specific requirement has far reaching implications on access to information and would result in a degraded user experience. We strongly disagree with this specific piece of the requirement and would recommend removing this for a number of reasons. ● The proposal appears to amount to a general monitoring obligation which is both extremely difficult to implement, and directly contradictory to Parliament's intention (as set out in the Summary above). ● 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 each child's unique profile of highlighted 'problematic content types'; ● A user may express dissatisfaction or negative sentiment towards a piece of content for a range of reasons - this can be based on not liking the creator who has uploaded the content, not liking the broad subject matter of the content, not liking the specific perspective provided on a subject, amongst others. Using a signal that suggests negative sentiment

	towards a specific piece of content is not helpful as a unilateral signal for determining what other content should also be downranked, and could result in a degraded user experience with a negative impact on children's access to certain 'categories' of information. ● It is unclear how this requirement will actually keep child users safer. Instead, this type of requirement would have a significant impact on access to information and will result in large amounts of false positives. Examples could include: ○ A user showing negative sentiment about content from a specific football team may actually be very interested in content from local rivals; ○ A user showing negative sentiment towards content criticising the 'green agenda' may be very interested in content relating to environmental issues as a whole; ○ A user expressing negative sentiment towards a new album from a band may be very interested in its previous works, or in other bands in the same genre. Suggested amendment PCU F3.4 - The provider should ensure that, where a child user has signalled negative sentiment towards a specific piece of regulated user generated content present on any of their recommender feeds, any downranking or removal is limited to that piece of regulated user-generated content. (This does not affect the application of Recommendations PCU F1 or PCU F2.) **Recommender systems** Ofcom states in relation to recommender systems that "our evidence indicates that content recommendation systems that are primarily optimised for user engagement can suggest content harmful to children (***CHC***), such as violent content, to children without them actively seeking this content out", which shows Ofcom's general presumption that a recommender system causes harm (Annex 6, table A1.1).

	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. They connect users to relevant, timely and high-quality information as we take the additional step of recommending authoritative videos to viewers on certain topics, such as those prone to misinformation. We rely on human evaluators, trained using publicly available guidelines, who assess the quality of information in each channel and video. We also rely on certified experts, such as medical doctors, when content involves health information. To decide if a video is authoritative, evaluators look at factors like the expertise and reputation of the speaker or channel, the main topic of the video, and whether the content delivers on its promise or achieves its goal. More specifically, when it comes to our younger users on YouTube, recommendations help connect them to **high-quality content and minimise the chances they’ll see low quality content.** For example, we raise high-quality kids and families content in recommendations that meet our quality principles. This is content that inspires curiosity, imagination, and creativity in younger viewers while helping older children consider issues such as diversity and inclusiveness. These high-quality content principles were developed in collaboration with our Youth and Families Advisory Committee, which consists of 13 experts in children’s media, child development, digital learning, and citizenship from a range of academic, non-profit, and clinical backgrounds. We consulted with the same experts to develop a companion set of low-quality content principles that impact channel performance. We use recommendations to reduce the spread of content that matches our low-quality principles — for instance, because it’s heavily commercial, pseudo-educational, or misleading. Ofcom’s conclusion about harm associated with recommender systems also does not take into account that different use cases for ranking content across different products will change the risk profile. For example, Photos organises a user’s own photos and videos into themes of meaningful moments, or “Memories”. Although this might technically be a means of

	ranking suggested content, all of the content is within that user's gallery (and therefore their account), rather than third party content, and so the risk of such content being harmful to the user that opted to store the photo within their account is drastically lower. We are unaware of any evidence that content recommendations (e.g. Memories) on Photos would increase the risk of child users encountering CHC or violent content. Suggested amendment Amend the risk profiles to reflect that recommender systems can be used to help decrease risk of harm, particularly for children. **GenAI** Similarly, Ofcom states in relation to GenAI that GenAI technologies present a risk to children online, given that children are early adopters of GenAI which can be used to facilitate the creation of CHC (Volume 3, para 6.12). Ofcom accordingly lists GenAI as a specific example of a risk that may occur where a service's growth strategy includes implementing new technologies (Annex 6, table A1.1). However, we note that GenAI is not necessarily inherently harmful to children, and the risks posed will need to be assessed in the context of the broader product and the feature that it's connected to. Assuming that GenAI is inherently riskier to children than other services fails to recognise that there may be nuances in the way in which GenAI is used on different products. Suggested amendment We recommend that Ofcom clarify that GenAI is not intrinsically harmful to children, and the risks that arise will need to be assessed in the context of the way in which GenAI is used on the service, and broader factors about the service itself.
User support (Section 21)	
53. Do you agree with the proposed user support measures to be included in the Children's Safety Codes? a) Please confirm which proposed measure your views relate to and	Confidential? **– Y (Partly)** **General comments** We agree with the intention behind the measures to provide greater support to child users, as a way of protecting children from encountering harmful content.

provide any arguments and supporting evidence. b) If you responded to our Illegal harms consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response.	We recommend that Ofcom include clear proportionality provisions, to reflect the fact that not all elements of the measures are appropriate for all types of U2U platforms. In particular, the requirement for services to provide children with the option to block and mute other users' accounts is likely to be difficult to implement for many services and would not reduce the risk of harm more than other similar features that have similar goals. The measure currently applies broadly to services that are medium/high risk for a broad range of harms, allow user connections and have a range of functionalities, including that of simply "posting content". Applying this prescriptive measure to such a broad range of services fails to take into account the different natures of the services within scope. We note that Google's response to Question 37 (20.1) of the Illegal Harms Consultation is also relevant to this response. **Suggested amendment - US2** We recommend that Ofcom scope Measure US2 more narrowly, to a more limited set of functionalities (e.g. not to any service that allows posting of content) and only to services that are at high (rather than medium or high) risk of the relevant harms. This would reflect the fact that the measure is not appropriate for all types of U2U platforms that fall within its scope, as currently drafted. [✂] **Suggested amendment – US4** We recommend that this requirement be less prescriptive, allowing online service providers to surface the required information outside of in-app features. This would enable providers to share information about effects and additional options when that information is most useful to the user: before they use a content restriction tool. This could be done through any manner of solution that may be appropriate for the risks and functions of the provider. **Suggested amendment – US5** We also recommend that a proportionality requirement is included in relation to Measure US5, given the difficulty in

	accurately identifying bullying content. (As set out in our response to Section 20.) Furthermore, we note that ECU 3.2 seems to state that subsequent obligations (ECU 3.6, 3.8, 4.0) only bite where a platform has "means of identifying when a user uploads, generates or shares one or more of the kinds of content for which it has a medium or high risk in accordance with sub-paragraph b)". We are generally concerned about the notion of platforms being asked to meet additional compliance obligations if they voluntarily take steps to protect children from harmful content. The rules seem to dis-incentivize platforms from seeking to identify these types of content. We recommend that Ofcom looks to remove this disincentive. Also, similar to US4, we recommend that US5 be less prescriptive for users reporting relevant content (PCU E3.4), allowing online service providers to surface the required information outside of in-app features (for example, in help centres). In-app notification may be particularly disproportionate for users reporting relevant content compared to users posting/sharing (PCU E3.5) or searching for (PCU E3.6) this content. These users would benefit from knowing the effects of reporting content before taking a reporting action, and are less likely to benefit from support than a user posting/sharing or searching for harmful content. **Suggested amendment – US6** These requirements are also overly prescriptive, and should be amended to give services discretion as to how the policy objectives should be met. For example, PCU E1.3 requires user support materials to include visual, audio-visual, or interactive elements. The support materials YouTube already offers are written to be accessible for our oldest and youngest audiences, as well as in between. Similarly, requiring providers to provide these materials in onboarding and search results (PCU E1.5 and PCU E1.6) is overly prescriptive, and does not consider what may be appropriate for each provider and their users.
Search features, functionalities and user support (Section 22)	

54. Do you agree with our proposals? Please provide underlying arguments and evidence to support your views. 55. Do you have additional evidence relating to children's use of search services and the impact of search functionalities on children's behaviour? 56. Are there additional steps that you take to protect children from harms as set out in the Act? a) If so, how effective are they? As referenced in the Overview of Codes, Section 13 and Section 17, the use of GenAI to facilitate search is an emerging development and there is currently limited evidence on how the use of GenAI in search services may affect the implementation of the safety measures as set out in this section. We welcome further evidence from stakeholders on the following questions and please provide arguments and evidence to support your views: 57. Do you consider that it is technically feasible to apply the proposed codes measures in respect of GenAI functionalities which are likely to perform or be integrated into search functions? Please provide arguments and evidence to support your views.	Confidential? – N We set out above in response to Section 18 our position in relation to restoring search content to its previous position.

Combined Impact Assessment (Section 23)	
58. Do you agree that our package of proposed measures is proportionate, taking into account the impact on children's safety online as well as the implications on different kinds of services?	Confidential? – Y / N
Statutory tests (Section 24)	
59. Do you agree that our proposals, in particular our proposed recommendations for the draft Children's Safety Codes, are appropriate in the light of the matters to which we must have regard? a) If not, please explain why.	Confidential? – Y / N
Annexes **Impact Assessments (Annex A14)**	
60. In relation to our equality impact assessment, do you agree that some of our proposals would have a positive impact on certain groups? 61. 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? a) 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	Confidential? – Y / N

use Welsh and treating Welsh no less favourably than English.	

Please complete this form in full and return to protectingchildren@ofcom.org.uk.