Our response

Executive summary

Google is supportive of effective content regulation and is aligned with Ofcom's goal of ensuring regulation helps keep users safe from bad actors while protecting the core benefits of online environments, including the ability of users to express lawful speech openly, access useful information and connect with one another. We believe in the power of the open internet and how it acts as a catalyst for innovation, economic growth, education and social well-being.

At Google, we have been working on this challenge for years, ensuring the right policies to protect our products and users, and using both computer science tools and human reviewers to identify and stop a range of online abuse, from "get rich quick" schemes to disinformation to child sexual abuse material. A mix of people and technology helps us identify illegal and harmful content and enforce our policies, and we continue to improve our practices and remain committed to transparency through regular updates to our Community Guidelines Enforcement Report.

We have not waited for new regulation before acting to keep our users safe. We are constantly improving and introducing new policy changes to support online safety and continuing to invest in technology to help us tackle illegal and harmful content at scale.

We recognise that tackling this problem is a shared responsibility, and we want to offer our thoughts to contribute constructively to the conversation.

Overall, we are broadly aligned with Ofcom's approach in the draft Codes and recognise the balance it has to strike on the specificity of the Codes given the scope and scale of the obligations and the variety of services to which they will apply. However, we consider the draft Codes would benefit from practical improvements to ensure Ofcom's policy objectives are met. This means doing so in a fair way so that we continue to raise the floor for services that may be newer to content safety measures and we don't inadvertently set a compliance ceiling for services that have long invested in providing responsible platforms.

Our suggestions are anchored on the following six themes:

Flexibility to allow for innovation in online safety: We appreciate that Ofcom is trying to set clear measures that are accomplishable for businesses of any size. However, the explicit instructions on how to meet measures documented in the draft code appear to leave little flexibility for the use of advanced capabilities to mitigate the same risk and benefit from the safe harbour protections set out in the OSA. Prescriptive provisions such as the use of 'fuzzy keywords' to detect fraudulent content, could unintentionally push platforms to adopt a "lowest common denominator" approach to comply with the law, rather than take a more sophisticated and effective approach that would not offer the same certainty of compliance. We would welcome it if the Codes could be framed more broadly in places, to ensure that a range of current and future technological solutions to compliance can benefit from the safe harbour provisions.

Proportionality of certain obligations: We would welcome further clarity from Ofcom on the proportionality of certain obligations rather than leaving it to platforms to rely on justifying 'alternative approaches' to the draft Codes. Also, where Ofcom gives precise examples, it should be clear that platforms are only required to implement those solutions where it would be proportionate to do so and where it directly relates to addressing a specific harm set out in the Act or in the platform's risk assessment. Some of the examples we include in our response include duties around appeals and complaints for downranking content and the lack of differentiation in the measures and risks that apply to different types of content across services.

Removal of illegal content and the risk of unintended consequences: We are concerned about the overly broad interpretation of 'reasonable grounds to infer' as it relates to a platform's duties on illegal content judgments. We consider the guidance should reflect that it is only 'reasonable to infer' that content is illegal if platforms reasonably believe that a court would also judge it so (even though a court order is not required). As currently drafted, the draft Codes may place a legal obligation on services to remove lawful content, effectively making a statement 'illegal' when made online, which would be legal offline. Further, the guidance suggests platforms should undertake a broad set of investigative requirements (including reviewing off-platform published information and previous user activity), impacting privacy and placing a disproportionate administrative burden on platforms. We believe the draft Codes should be updated to set some clear limits on the information that platforms are expected to reasonably consider as part of these judgments and more emphasis should be placed on key flaggers submitting relevant information.

Approach to risk and risk assessments: The draft risk assessment guidance does not clearly distinguish between inherent (i.e. before risk mitigation measures) and residual (i.e. after risk mitigation measures) risk, which means that the adequacy of existing compliance measures may not be taken into account when the draft Codes recommend further compliance measures. Further, we have concerns around the definition of what constitutes 'significant' change to a product or service that would trigger an update to a risk assessment. In our detailed response, we provide some practical suggestions that would meet Ofcom's policy objectives whilst ensuring that platforms have certainty as to when they need to update their assessments. We recommend that proposed platform changes to recommender systems are assessed and documented under the same parameters as all other platform changes.

On a related note, we also encourage Ofcom to reconsider how it positions recommender systems as largely a risk vector. Recommendations don't just help connect viewers with content that uniquely inspires, informs and entertains them, they play an important role in how we maintain a responsible platform. On YouTube, recommendations compliment 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. In addition, we have used recommendations to limit low-quality content from being widely viewed since 2011, when we built classifiers to identify videos that were racy or violent and prevented them from being recommended and to improve user experience. Since 2019, YouTube has worked aggressively to reduce recommendations of borderline content and harmful misinformation. The more "borderline" a video, the less frequently it is recommended. We would welcome Ofcom's updated Codes to reflect the positive role recommendations can play in keeping users safe.

Clarity on general monitoring: We would encourage Ofcom to explicitly set out that none of the duties in the draft code would oblige platforms to proactively monitor the service for illegal content (most notably for fraud and CSAM). This was certainly not the intention during the legislative phase, as confirmed from the despatch box, and we understand that is not Ofcom's policy intent either. However, we believe further clarity from Ofcom would add regulatory certainty and reduce the risk of significant over removal of legal content.

Implementation timelines: We recognise that Ofcom faces pressure to ensure rapid implementation of the OSA. However, it is important to ensure that the industry has time to adequately digest, respond and prepare for compliance. Where compliance with the Codes requires platforms to undertake significant systems and process changes, Ofcom should take into account that this process can only start once the provisions have been finalised, and therefore allow a more realistic implementation period (of 9-12 months).

As part of our response, we have highlighted the specific provisions of the draft Codes where clarifications or amendments could be made, to help Ofcom better understand the points we make below.

Should you and your team have any questions about our response, we are more than happy to meet Ofcom and discuss.

Question (Volume 2)	Your response
Question 1 (6.1): Do you have any comments on Ofcom's assessment of the causes and impacts of online harms? Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.	**Confidential: N** We note that the causes and impacts of online harms as described in Ofcom's guidance are necessarily general. However, the causes of harm in any individual case are likely to be complex and multifactorial, including offline/ real world experiences, such that something that causes harm to one person may not cause harm to another. Equally, the unique profile of the user (and their specific attributes and characteristics) are unlikely to be ascertainable by the service. Volume 2 describes the evidence that forms the basis of Ofcom's Register of Risks. Ofcom's Risk Profiles, which services are required to take into account as part of their risk assessments, rely on this evidence base. Given the centrality of the evidence to the risk assessment process, it is critically important that this evidence base is reliable, robust, ethical, independent and methodologically sound, as Ofcom itself has acknowledged (Vol 2, 5.10 read with FN 5). This is also consistent with Ofcom's duties as a public authority to ensure that it proceeds on the basis of robust evidence, as it formulates codes of conduct and guidance with which in-scope companies will be expected to comply. However, we are concerned that there are significant gaps in the evidence base upon which Ofcom relies in places (e.g. Vol 2, 5.17). We are also concerned that certain of the research cited in Volume 2 does not meet Ofcom's own standards for evidence being, at times, out of date, no longer accurately reflective of the market or harms described or lacking appropriate methodology/peer review (Vol 2, 5.17). Ofcom should ensure that the evidence base for the Codes and Guidance meets its own evidential standards and provides a sufficiently robust and reliable basis for regulation. In this context, it should also clearly identify provisions of the Codes and/or Guidance for which insufficiently robust evidence is available; this will be relevant to the weight which in-scope companies, and we assume Ofcom itself, can be expected to place on those provisions once the codes and guidance are in force. We would also welcome clarification about how and why Ofcom has selected the sources in accordance with the

Question (Volume 2)	Your response
	requirements for evidence that it sets out in chapter 5 of Volume 2. Please also see our response to **question 2** where we provide a more detailed response on the links Ofcom has made between risk factors and different kinds of illegal harm.
Question 2 (6.2): Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer.	**Confidential: N** The Act requires Ofcom to prepare risk profiles and we recognise the work that Ofcom has undertaken to gather evidence and assess the causes and impacts of illegal online harms. In outlining its assessment, we recognise the need for Ofcom to draw generalisations about the links between risk factors and different kinds of illegal harm. However, as recognised by Ofcom (at Vol 2 para 5.26), not all characteristics are inherently harmful, and some functionalities which Ofcom has identified as risk factors that may be linked to particular kinds of illegal harms can indeed be beneficial for consumers (at Vol 2 para 6.11). We believe Ofcom should explicitly recognise that characteristics do not necessarily correlate to an increased risk of harm in the context of every service that has that characteristic. We do not comment on every conclusion or evidence relied upon in Volume 2, but would like to highlight below the three areas where this is a particular challenge, namely: the role of recommender systems, end-to-end encryption and advertising as a business model. **<u>Recommender systems</u>** Ofcom's guidance states that although "*recommender systems deliver content to users on your service that they may find interesting, they can also lead to a risk of harm*".[1] **Google is concerned that the Codes currently over index on the potential harm caused by illegal or harmful content with the function of recommender systems.** Recommendations don't just help connect viewers with content that uniquely inspires, informs and entertains them, they play an important role in how we maintain a responsible platform. <u>YouTube:</u>

[1] Annex 5, page 60

Question (Volume 2)	Your response
	As a video sharing platform, recommendations compliment 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. The more authoritative a video, the more it is promoted in recommendations. In addition, we've used recommendations to limit low-quality content from being widely viewed since 2011, when we built classifiers to identify videos that were racy or violent and prevented them from being recommended, and to improve user experience. Since 2019, YouTube has worked aggressively to reduce recommendations of borderline content and harmful misinformation. The more "borderline" a video, the less frequently it is recommended. Connecting viewers to high-quality information and minimising the chances they'll see problematic content is not just important from a platform safety perspective, it is also paramount to our goal of recommending content that delivers value. These efforts complement the work done by our robust Community Guidelines, by allowing content that some may find objectionable to remain visible and accessible to users on the platform who wish to find and view it, and are critical to our responsibility efforts. The basic operation of many modern online services involves constant updating of content and features optimised for users. Algorithms and recommendations are constantly adapting to signals for our users, our authoritativeness classifiers, user surveys, and other techniques to improve our products for our users. Our recommendation systems are constantly evolving, learning every day from over 80 billion pieces of information we call signals. In addition, YouTube provides users with transparency and control over recommendations and continues

Question (Volume 2)	Your response
	to invest in improving users' experience to maximise the value we bring to users. For example, users have several options available to them that are not based on profiling, such as the Explore tab, topic channels, Subscriptions tab and Channels pages. In addition to exploring content not based on profiling, we provide information to users about how they can manage their recommendations. We've built controls that help users decide how much data they want to provide. Users can decide how they share their watch and search history data with us. Users can pause, edit, or delete their YouTube watch and search history whenever they want. We are transparent about how users can access these controls and the signals that inform our recommendations. We also note that this overly-negative characterisation of recommendations has additional impacts on other areas of the Codes. For example, we do not agree with the position that changes to recommender systems require risk assessments even when minor, and not having a significant impact on user harm. The bar should be 'significant change' in both instances - see also **questions** [**7, 9 and 34**] Search: Ofcom recognises that search services provide "significant benefits to individuals and society" in Volume 2 at paragraph 6T 15. On Google Search, we seek to provide the most relevant and authoritative results possible. We use ranking algorithms to ensure we are meeting users' expectations of surfacing relevant and high quality sources, as well as minimising low quality or harmful content from appearing prominently in search features or search results, where users are not actively seeking out such content. The design of these systems is our greatest defence against harmful content and other types of low quality information, and it is work that we've been investing in for many years. Some of Ofcom's conclusions about the risks of harm on search services appear to be contradictory or insufficiently substantiated. For example, Vol 2 6T.22 states that "*using search services is an effective way for users to access illegal content*" *but also that "there is limited evidence on the volume of illegal content directly accessible via search services*" (Vol 2, para

Question (Volume 2)	Your response
	6T.22). Equally, Ofcom states that the mechanisms by which illegal content can manifest itself may be different on search services compared to U2U services, but also that "*the impact on individuals is comparable*" (Vol 2, para 6T.16). There is no evidence to support this conclusion, or the guidance that readers should refer to the evidence in the user-to-user chapters to understand how harms manifest on search services. We would like to see Ofcom reconsider these statements and explain its logic behind them (Vol 2, para 6T.16). Other services: Ofcom's conclusion about harm associated with recommender systems does not take into account different use cases for ranking content across different products. For example, Photos organises 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 it is difficult to envisage how the use of Memories could increase the risk of user harm. Ofcom considers that there is a risk that content recommender systems "*inadvertently amplify illegal content to a wide set of users who may otherwise not organically come across this content. Our evidence, for example, indicates that if not properly tested and deployed, content recommendation systems may amplify hateful content if they are optimised for user engagement.*"[2] However, this conclusion does not bear weight in every context, since there is no evidence that content recommendations (e.g. memories) on Photos would increase the risk of users encountering illegal content generally, or hateful content particularly. **End-to-end encryption ("E2EE")** Ofcom's guidance states that encryption is likely to lead to increased risk of a range of offences, including fraud and financial services related offences (pg. 3). However, it is important to note that in certain settings, and if used correctly, E2EE can also enhance user protection against fraud and identity theft: for example, the FCA recommends encryption of

[2] Annex 5

Question (Volume 2)	Your response
	customer data for firms in financial services, to enhance data security. Rich Communication Services (RCS) chats are a modern industry standard between telecommunications operators and mobile phone carriers, which boost consumer confidence and reduce fraud through the implementation of a verified sender system. As noted by the ICO during the Online Safety Act's Committee Stage,[3] E2EE supports the security and privacy of online communication and we agree that Ofcom's online safety regime should not inadvertently trade one perceived risk for another. We believe Ofcom's draft Codes should reflect the positive role that such technologies can play. **Advertising** Paragraph 6.5, Volume 2 suggests that advertising should be considered as part of the Risk Register as a characteristic of a service (as part of that service's business model). Advertising is considered in this context in relation to each offence (e.g. 6B.70-73 suggests that advertising models may in principle reduce the risk of harm in relation to terrorism; and 6C.192 suggests that it could increase the risk of CSAM). However, requiring services to consider the impact of advertising in relation to all priority offences is disproportionate and inconsistent with the deliberate limits placed on the application of the online safety regime to advertising by the Act itself. The Act requires only Category 1 and Category 2A services to consider ads, specifically in the context of the fraud offences set out in s.40 (since they are required to use proportionate systems and processes to prevent users from encountering fraudulent ads). Bringing advertising within scope of the Act in this limited way was a considered policy choice driven by the desire not to duplicate parallel efforts to reform the regulatory framework for paid-for online advertising, such as through the Online Advertising Programme. There is no standalone requirement under the Act on advertising services to mitigate the risk of users encountering all forms of illegal content.

[3] ICO Director of Technology and Innovation, Stephen Almond, 26 May 2023, page 82 https://publications.parliament.uk/pa/bills/cbill/58-03/0004/PBC004_OnlineSafety_1st17th_Compilation_29_06_2022.pdf#page=82

Question (Volume 2)	Your response
	We would therefore suggest that only Category 1 and Category 2A services are required to consider advertising as part of their risk assessments, and only in relation to fraud.

Question (Volume 3)	Your response
Governance and accountability	
Question 3 (8.1): Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice? Please provide underlying arguments and evidence of efficacy or risks to support your view.	**Confidential: N** Annex 7, A3.4 requires services to name a person accountable to the most senior governance body for compliance with illegal content safety duties and the reporting and complaints duties. Given the flexibility that the Act envisages for services to designate a responsible person *at the point of enforcement action or an information notice being issued*, we would make the following recommendations in relation to this provision: ● The Codes should recognise the complexity of large and multi service platforms and allow them to name an accountable function for these purposes.[4] This will allow services to nominate more than one person and better ensure this is aligned to the specific safety aspect in question. Nominating more than one individual would be consistent with similar regimes in other contexts, such as the Senior Managers regime in financial services. ● Platforms should expect to have some discretion over which individual(s) is/are the appropriate senior manager(s) to be named in relation to an information notice or enforcement activity, and it will not necessarily be the same person as whomever is accountable to the most senior governance body pursuant to the Code of Practice.

[4] See also A3.4, Annex 8 in relation to search services.

Question (Volume 3)	Your response
Question 4 (8.2): Do you agree with the types of services that we propose the governance and accountability measures should apply to?	**Confidential: N** See responses to **Q14 and 15** below regarding the definition of large and multi-risk services**.**
Question 5 (8.3): Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party?	**Confidential: N** Since Ofcom is an evidence-based regulator we would expect that it would consult, and provide evidence, on the benefits of introducing independent third party audits, if it were minded to pursue this as an option. In our view, such an obligation would risk imposing a disproportionate burden on services given the potential resources needed to fund and facilitate an audit. Depending on the precise scope of the audit and proposed audit processes any additional risks to services, users and third parties will also need to be given due consideration. Further, we are not aware of any evidence that an external regulatory audit increases efficacy; rather, it diverts significant resources -- in terms of time spent by internal stakeholders, as well as cost -- from actually mitigating the risk of illegal content. For example, in the context of DSA, it is generally the same teams and stakeholders involved in developing mitigation measures resulting from a risk assessment, and preparing for the next risk assessment cycle, but also being required to support the audit process. Equally, our current estimates of the costs associated with an independent audit are anticipated to exceed $10 million. There's also no evidential basis to suggest it is necessary, effective, or proportionate. This is especially true in the abstract insofar as the scope of such a potential audit and the standards against which it would be judged are unclear. On the contrary, there is evidence to suggest[5]

[5] See for example: https://www.nytimes.com/2023/12/28/us/migrant-child-labor-audits.html

Question (Volume 3)	Your response
	that audits in pre-existing regimes are inefficient and often fail to achieve their stated purpose.
Question 6(8.4): Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to tie remuneration for senior managers to positive online safety outcomes?	**Confidential: N** We recognise the importance of structural incentives. However, a measure which seeks to tie remuneration for senior managers to positive online safety outcomes would be neither workable in practice nor consistent with the scheme of the Act. This sort of measure is not workable in practice because there is no principled basis to compare "online safety outcomes" consistently and reliably between services. Such a measure is also not consistent with the scheme of the Act and risks improperly shifting focus away from the adequacy and appropriateness of systems and processes. The premise of the Act is that different services have different inherent risk levels. And while all services are required to implement appropriate measures to mitigate risk, the existence of unmitigated (or residual) risk is both unavoidable and recognised by the scheme of the Act. The residual risk profile of services will differ based on inherent attributes of the service in addition to matters senior managers can reasonably be expected to control. During the passage of the Online Safety Act, parliamentarians and campaigners were clear about balancing the importance of safety concerns with the rights to freedom of expression. If senior managers are incentivised by the removal or sanitisation of content, there is also a real risk of unintended consequences, including the over-removal of lawful content and removing adult users' access to legitimate free speech. This in turn may shape other, less democratic states' approach to content moderation and could encourage autocratic regimes to draft legislation and regulate in a way that undermines human rights and values. The UK Government has committed to developing the UK into a 'tech superpower'. For all tech businesses, from start ups, to established global companies, the impact of regulation on individual staff and talent acquisition/

Question (Volume 3)	Your response
	retention will be an important factor in location decisions. Therefore, it is important to the wider economy that such measures are considered with due proportionality.
Service's Risk Assessments	
Question 7 (9.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: N** 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 Act. 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) assessed the impact of the policy change, and whether it has addressed the relevant harm. Ofcom's guidance is therefore 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 Risk Assessment Guidance may be helpful for some businesses, in general the approach is overly prescriptive and it will limit the ability of more established businesses, such as Google, to align risk assessment processes under the Act with existing organisational risk management processes. In addition to imposing significant compliance costs, the Guidance is, in places, disproportionate and lacks flexibility (as explained further below). Whilst 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 non-exhaustively some specific aspects of the risk assessment proposals where we

Question (Volume 3)	Your response
	believe surgical improvements could be made to the Codes in order to better deliver the policy intent. **“Significant change” meaning** 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 Act. While parts of the Guidance (e.g. at Vol 3, para 9.138) recognise that it would not be proportionate to capture routine changes and upgrades, this principle is not reflected in the operational parts of the Guidance. Importantly, if the obligation to risk assess a change is inappropriately triggered there could be unintended consequences. This is a particular concern if the threshold for a “significant change” is set too low, as it is now, as it could discourage services from regularly making product improvements, including those that reduce risk, because of compliance costs associated with risk assessment requirements. To illustrate this point, in 2022, Google Search had over 4,000 ‘launches’ which were underpinned by over 13,000 live experiments, nearly 900,000 search quality tests and nearly 150,000 side-by-side experiments. It would not be workable or proportionate to expect Search to update its risk assessment for each launch or update.[6] Even if higher thresholds are included, a broad definition of “significant change” risks stifling innovation - particularly for larger services that have greater resources to use for innovation. The Guidance’s general focus on the size of a service’s user base as relevant to whether a change is significant has no basis in the Act. Whilst, in principle, there may be some circumstances in which the significance of a change may be amplified by the size of the service, we do not consider that this should be the main determining factor. In fact, Ofcom itself

[6] Please see: https://www.google.com/search/howsearchworks/how-search-works/rigorous-testing/

Question (Volume 3)	Your response
	recognises that in some instances the number of users may be a weak indicator of risk level (Vol 3 para 9.62). The size of a service's user base has no independent correlation to a service's risk profile. A neutral change (e.g. to the indent of bullets in a list) is not rendered more risky because the service has a large user base. Noting the Guidance (at [A5.133]) contemplates that many significant changes will require a service to carry out an entirely new risk assessment, rather than one directed only to the change, it would be an entirely disproportionate for large services to undertake a new risk assessment each time they seek to make a change to the service. This approach may have the effect that large services are in practice only able to make bundled changes on a periodic basis (e.g. yearly). Not only would this disproportionality impact services' commercial interests, but for the reasons outlined above would likely result in less frequent changes directed to improving the service and reducing risk. In this way, the Guidance in this area risks undermining the goals of the regime.. There is also an inconsistent approach to a materiality threshold in the examples and Table 13 and paragraph A5.135 in the draft Guidance. This has the effect that it appears as though a "significant change" could include such minor changes that are unlikely to have any material impact on the availability of certain content on the service (e.g. an amendment to a content moderation policy that clarifies a prohibition on images of deadly weapons includes assault rifles, or a change in the location or design of a "report" or "react" button). <u>Suggested amendment</u> ● Revise the Guidance to make clear that the size of a service's user base is not independently sufficient to render a change significant. For example, it would be helpful to remove, or add a qualifier to, the trigger a "*proposed change which impacts a substantial proportion of a service's user base or changes the kind of users you expect to see on your service*" from Table 13 on page 48 of

Question (Volume 3)	Your response
	Annex 5. “Significant” is already linked to the incidence/likelihood of harm because Table 13 states that a change will be significant if it "alters the risk factors which you identified in your last risk assessment." ● Ofcom may also consider improving the comprehensibility of the guidance by re-formatting the criteria identified in Table 13 as a list of considerations rather than in the current tiered structure. **Use of user data** 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 with how the balance between safety and privacy is being approached in the Guidance, in particular regarding the potential requirement to collect and process personal data in connection with illegal content risk assessments. In the Guidance, Ofcom indicates that services should consider user base demographics, and the vulnerability of users in relation to gender and protected characteristics, such as age, race, ethnicity, sexuality, sexual identity, religion and disability when considering the “user base” factor for illegal content risk assessments. In addition, Ofcom clearly envisages that methods for the collection of this data include processing of personal data and could include data derived from age verification processes, behaviour identification and user profiling technology (Annex 5 p.39 and Vol 3 p.73). However, it should be recognised that it would be rare for a service to have this level of information about their user base demographics or a particular user’s characteristics, and nor can this information be reliably inferred from user behaviour. For example, a user searching for information

Question (Volume 3)	Your response
	about Judaism or a Jewish festival, is not necessarily Jewish themselves. Even if these kinds of data could be collected in practice, the Guidance should also take into account that any new collection of personal data would need to be proportionate to the related benefits of online safety. In particular, the type of information envisaged in the Guidance above (e.g. ethnicity and religious beliefs) constitutes special category data. The processing of special category personal data would need to meet additional conditions, and on the basis of the current Guidance it is difficult to see on what basis it would be lawful. Instead, Ofcom should allow services to use external publicly available data or insights from external experts and civil society organisations to inform their consideration of vulnerable users. As recognised by Ofcom, services will need to consider carefully whether the processing of this data for the purposes of the risk assessment, to the extent that it constitutes personal data and/or special category personal data, is compliant with that service obligations under the UK GDPR and Data Protection Act 2018. In addition to being satisfied that the processing is proportionate, and compliant with the purpose limitation and transparency principles, services would need a lawful basis under the UK GDPR to process personal data. Our understanding is that it is Ofcom's intention that it is for services to decide whether they should or can lawfully process any personal data in connection with risk assessments (including identifying an appropriate lawful basis). However, Ofcom also has a duty under s99(4) of the OSA to consult with the Information Commissioner Office (ICO) before producing guidance on risk assessments. We consider that these two regulators should first look to avoid confusion between the application of the UK GDPR and the OSA and then appropriately update the risk assessment guidance. Suggested amendment

Question (Volume 3)	Your response
	- Revise the Guidance to make it clear that: - services are responsible for determining which personal data, if any, they will process as part of the evidence for the risk assessment; - any new collection of personal data would need to be weighed against the privacy risk and the additional intrusion into privacy be proportionate to the benefits of online safety; and - there is no requirement to process special category personal data. - services may instead use external, publicly available data about services or insights from external experts and civil society organisations to inform their consideration of vulnerable groups. - Consistent with Ofcom's duty under s99(4) OSA, it must consult with the Information Commissioner before producing guidance on risk assessments. **Scalability/proportionality** Ofcom recognises that there is a desire for its approach to risk assessments to be scalable - see, for example, Volume 3, 9.24(e). However, we are concerned that Ofcom's approach may not be scalable in practice and, in places, could be disproportionate to the risk of harm (in particular, the underlying assumption that the size of a service is directly proportional to the risk). This is particularly evident, for example, at A5.109 with the assumption that larger services are likely to have the resources to include enhanced inputs, without reference to the risk. And again at Table 6, where there is an underlying assumption that the impact of harm is dependent on the size of the service. Ofcom's assumptions regarding evidence are an area of concern. This includes the assumption that core data will be readily available to services (Vol 3 9.107). Some data, such as quantitative data surrounding user complaints and content moderation are not necessarily currently collated

Question (Volume 3)	Your response
	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 it is not proportionate to the risk to collect that data in a detailed and granular form for all services. The form of the data collected should be proportionate to the risk profile of the service. We also question the usefulness of the insights that may be derived from that data at the level of granularity expected by Ofcom and the reliability of the data, given that users do not necessarily give reasons for a complaint or flag with accuracy or precision. These concerns about proportionality are particularly relevant for smaller services. We would like to see Ofcom offer services greater flexibility in determining which evidence will be appropriate to the risks identified on their services. Suggested amendments: Remove the assumption that larger services are required to use enhanced inputs irrespective of risk.[7] Remove the assumption that the impact of harm is dependent on the size of the service.[8] Revise the Guidance to give services greater flexibility in determining which evidence will be appropriate to the risks identified on their service, particularly where there may be an impact on vulnerable user groups. Revise the Guidance to clarify that: ● the Guidance is not prescriptive regarding the form and types of evidence that may be used for risk assessments, by deleting paragraph 9.107 of Volume 3 and including a statement in Annex 5 that services should determine which evidence will be appropriate to the risks identified on their services; and

[7] E.g. At A5.16, A5.109, A5.117, Annex 5

[8] E.g. At Table 5, Table 13, Annex 5

Question (Volume 3)	Your response
	● Ofcom’s assumptions regarding evidence will not apply in all respects to all services by including a statement to this effect in Annex 5 and the “Core and enhanced evidence inputs” section of Volume 3. **Inherent vs. residual risk** Consistent with requirements of the Act, the guidance in Volume 3 appears to focus on residual risk (i.e. risk after mitigation measures) rather than inherent risk (i.e. risk before mitigation measures). However, in some places this could be made clearer. In particular, certain parts of the guidance, especially relating to the “impact criteria” (e.g., number of users, gravity of harm) uses language sometimes associated with inherent risk. The draft Risk Assessment Guidance should be amended to clarify that the risk assessment obligation under the Act is directed toward residual risk and that, in assessing residual risk, services should have regard to the residual impact risk and the residual likelihood risk. Suggested amendment We would encourage Ofcom to use the language of “inherent” and “residual” risk to align the draft Risk Assessment guidance with existing risk management best practice. While this language does not appear in the Act, the Act is clearly directed to residual risk (since it requires services to “effectively mitigate and manage” risk). In particular, to reflect the Act’s objective, we would suggest the following changes to the Risk Assessment Guidance (Annex 5): ● In section 2.3, Ofcom could explicitly clarify that the risk level that is assigned relates to the residual risk; ● In Table 4 on page 19, the language that encourages services to assess whether “there are systems or processes already in place that reduce the risk of harm”, when assessing “likelihood of

Question (Volume 3)	Your response
	harm", could be repeated in Table 5 on page 20/21, which addresses the list of what services should consider when assessing the impact of harm; and ● In Table 6, the "Risk level table", Ofcom could clarify that the risk being assessed is the residual risk.
Question 8 (9.2): Do you think the four-step risk assessment process and the Risk Profiles are useful models to help services navigate and comply with their wider obligations under the Act?	**Confidential: N** Ofcom states that its intention is to reflect best practice and current standards in risk management (for example, Volume 3, [9.25]). While we appreciate that Ofcom's four-step approach to risk assessments may be an example of good risk management practice, it also notes that large companies, such as Google, will likely have sophisticated risk management practices in place already. It is disproportionate not to contemplate some flexibility within the Guidance for organisations with existing governance structures and escalation routines, so that management has a clear view of risks across the organisation. We consider that Ofcom should acknowledge that there may be some variation in risk management practices depending on the nature of the company and its services and that these practices may be appropriate alternatives. For example, Ofcom provides guiding questions on assessing impact in Table 5 of the draft Service Risk Assessment Guidance. Ofcom should clarify that these are merely examples and that there is scope for other enterprise risk frameworks or human rights assessment best practices to be applied. As drafted, the risk profiles appear to take into account only inherent risk. Mitigation measures and benefits inherent to the relevant risk factor are not considered, which may skew the conclusions in the overall assessment. Please see our response to **question 9,** which expands on our concerns with this approach.

Question (Volume 3)	Your response
Question 9 (9.3): Are the Risk Profiles sufficiently clear and do you think the information provided on risk factors will help you understand the risks on your service?[9]	**Confidential: N** We are concerned that the risk profiles will be difficult to apply in practice. In particular: ● The risk factors have low thresholds - often, the mere existence of a feature/ability makes the risk factor applicable. For example, the threshold for discussion forums or chat room services appears to be **allowing** users to send or post messages (Annex 5, p 55). All discussion forums and chat rooms, by definition, permit the sending or posting of messages and, by Ofcom's logic, would have the relevant risk factor applied. In addition, where risk factor definitions include language such as "typically" as a threshold (for example, the definition of "messaging service" in Annex 5, p53), it will be difficult for services to understand if the risk factor applies to the service. ● For services that allow child users, Ofcom has provided a different definition of the threshold than provided in the Act in the context of children's access assessments. At Figure 3 (page 54) Table 14 (page 56) and Table 15 (page 64) the test is whether the service "allows child users", whereas in ss35 and 37 OSA there is a more detailed explanation of what is meant by "likely to be accessed by children". While there may not be much practical difference in outcome, it is unclear if the policy intent was to create a separate threshold. If not, it might be helpful if the Guidance could align with the statutory definition, to avoid disparity between the Guidance and the legislation. ● For some risk factors, it is unclear if Ofcom is seeking to repeat the requirements of ss.9(5) (e.g. user base, functionality, business model) or add considerations. This may lead to "double counting" in assessments.

[9] If you have comments or input related to the links between different kinds of illegal harm and risk factors, please refer to Volume 2: Chapter 5 Summary of the causes and impacts of online harm).

Question (Volume 3)	Your response
	We also have concerns around the appropriateness of the "service type" risk factors: ● They appear to be combinations of the functionality risk factors. It is unclear why it is useful to analyse both and how the results of each interact. We think that it would be more logical to break a service down into its features and analyse those features regardless of the service type assigned in the abstract. The social media risk factor, in particular, is defined very broadly, and covers all risks. This makes the relevance of the other risk factors unclear. ● For some service types, Ofcom appears to assume that certain service types have certain features or functionality. For example, Ofcom appears to assume that file storage services allow embedding on third party services (Vol 1, 3.45) and that messaging services support closed groups (Vol 1, 3.31). No explanation is provided for how risks should be adjusted if the feature is not present on the service. The guidance also does not reflect circumstances where a feature or functionality is ancillary or only a minor aspect of the service (such as a minor user-to-user feature on a navigational tool), and we consider it is important that, for a functionality to have the consequences stated in the risk register, it should be a significant feature of the service (in order to avoid incorrect assessments of harm). Elsewhere, Ofcom recognises that it is difficult to generalise and categorise services and therefore for services to understand which risk factors apply to them (Vol 4, 9.80(b)). ● The approach taken in the Guidance is inconsistent with the approach explained in Volume 3 regarding looking at factors rather than taking a "service type" approach to risks. Ofcom specifically acknowledges that *"services that may fall into the same type can have very different risks"* (Vol 3 para 9.80(a)) yet in the Guidance makes assumptions

Question (Volume 3)	Your response
	about the risks that will appear on certain services (Annex 5 page 55). Volume 2 recognises that many of the aspects that are captured in the risk profiles can deliver benefits (for example, at 6.11 generally and 6W.22 on recommender systems). However, the risk profiles at Annex 5, which are based on Volume 2 do not reflect this and only consider the risk of harm. We would welcome Ofcom carrying through the recognition from Volume 2, that "some of the risk factors can also be beneficial to users" (Vol 2 6.11) to the risk profiles in Annex 5. Suggested amendments ● Thresholds should be increased and be clearly-defined. For example, thresholds may be linked to a certain percentage of users on the service using the feature or function. ● For services that allow child users, Ofcom should align the Guidance with the statutory threshold, to avoid disparity between the Guidance and the legislation. ● We would welcome clarification from Ofcom on whether it is seeking to repeat the requirements of ss.9(5) (e.g. user base, functionality, business model) or add considerations. ● "Service type" risk factors should be removed. ● Where risk factors are linked to a particular feature or functionality, that functionality must be a significant (rather than ancillary) aspect of the service. Ofcom should include in the risk factors the recognition from Volume 2, that "some of the risk factors can also be beneficial to users" (Vol 2 6.11) to the risk profiles in Annex 5.
Record keeping and review guidance	

Question (Volume 3)	Your response
Question 10 (10.1): Do you have any comments on our draft record keeping and review guidance?	**Confidential: N** Although this guidance does not address children's access assessments, we would welcome future guidance allowing services to self-certify that a service is likely to be accessed by children and that the "child user condition" is met, without having to identify child user counts or conduct a formal children's access assessment. We consider that this would reduce the regulatory burden on both the services and Ofcom, in circumstances where there is no disagreement over whether the child safety duties apply to that service.
Question 11 (10.2): Do you agree with our proposal not to exercise our power to exempt specified descriptions of services from the record keeping and review duty for the moment?	**Confidential: N** For some smaller or very low risk products, such as Kaggle (a forum for data scientists to exchange coding tips), it would be disproportionate to undertake a compliance review at least annually (which Ofcom suggests is the minimum required frequency for compliance reviews at page 87 of Volume 3). In line with our comments at **Question 7** under "scalability/ proportionality", we would like to see Ofcom offer services greater flexibility in determining when it is appropriate to conduct compliance reviews. Similarly, it could be disproportionate to require all services to update their risk assessments annually (see page 39 of Volume 3). We would like to see risk assessments conducted at the frequency appropriate to the risks identified on the relevant service. This may mean that services only update their risk assessments when there has been a significant change.

Question (Volume 4)	Your response
Codes of Practice - general	

Question (Volume 4)	Your response
Question 12 (11.1): Do you have any comments on our overarching approach to developing our illegal content Codes of Practice?	**Confidential: N** We appreciate Ofcom's efficient development of the extensive illegal content Codes and the steady progress it is making on the implementation of the Online Safety regime. Our primary recommendation is to simplify the Codes so they are more principles-based rather than being overly prescriptive or mandating specific types of technologies **(see further Q16 below)**. The core focus for platforms should be the duty to mitigate harm from illegal content. Provided the measures are sufficiently effective, platforms themselves are best placed to decide on the most appropriate technological measures to implement and have the flexibility to adjust these over time to address threats as they emerge on their service. **We recommend Ofcom avoid mandating specific technological solutions or, where this is not possible, expressly note in the Codes that these measures are only illustrative examples of how the safety objective can be achieved.** We note that the consultation documentation spans more than 1700 pages of complex and detailed material (including evidence and research relied on by Ofcom to support its preliminary conclusions). This is a large volume of material, which services have needed to digest in a short time frame, alongside other parallel Ofcom consultations, research reports and requests for information. It would therefore be helpful if Ofcom could space its documents in a way that would allow stakeholders of all sizes to digest the information and participate in the consultation process.
Question 13 (11.2): Do you agree that in general we should apply the most onerous measures in our Codes only to services which are large and/or medium or high risk?	**Confidential: N** We note Ofcom's general principle of targeting platforms with a large reach, and in some instances demonstrating a medium or high risk of harm. We also share Ofcom's policy intent of reducing harms across services of all sizes. However, we would recommend Ofcom reconsiders the details of the designation of measures: ● We would suggest that the more onerous obligations should not apply where a service is only "large" (without also being at high risk of particular content) based on the fact that high user counts is not necessarily an indicator of risk to users.

Question (Volume 4)	Your response
	On the contrary, many large platforms have long-established risk mitigations in place that have been developed over the course of several years. Furthermore, high user reach is not necessarily indicative of a service being high risk, either due to the primary use of the product (e.g. navigation) and/or due to limited user-to-user functionalities. Smaller platforms also pose risks to users and if obligations remain tied to size of service alone, it may lead to pushing harmful content onto smaller, but higher risk services that are less regulated without the same incentives to manage risk. ● We would also encourage Ofcom to limit more onerous obligations to large services at **high** risk of harm, to ensure meaningful differentiation between obligations for services of medium risk or high risk. In the current Codes, most measures apply to services which is assessed as being medium or high risk for one or more specific kinds of harm (ie. a "specific risk" or "multi risk" service). It would be beneficial to limit the more onerous measures to services assessed as being "high risk" for one or more kinds of harm only, as this would increase the incentives on services to reduce the risk of harm from high to medium or below. Furthermore, the more onerous obligations should only apply where the service is high risk based on **residual risk** (rather than inherent risk) following risk mitigations, which may relate to the design of the service or the policies and processes they have in place to tackle various harms. Suggested amendment ● In relation to measures that apply primarily because of the size of a service (in particular, 3A, 3C, 3D, 3E, 3F, 3G, 4B, 4C, 4D, 4E and 4F, which apply to large services, regardless of whether any medium or high risks have been identified), we recommend that Ofcom remove the requirement for these services to apply to "large services", and only require them to apply to "large" and "multi-risk" services. ● Alteration of the definition of a "multi-risk service" to focus on a) residual risk, as described above; and b) services assessed as posing one or more "high" specific risks of illegal harm.

Question (Volume 4)	Your response
Question 14 (11.3): Do you agree with our definition of large services?	**CONFIDENTIAL: Y (partly)** The bar for “large services” is set too low, as it refers to services with more than 7 million monthly users. Ofcom suggests that this is consistent with the DSA approach to VLOPs (i.e. based on roughly 10% of the population). However, the DSA also uses a functionality test in practice given that it only applies to search engines and online platforms (which have to allow for the public dissemination of content that is not a minor and ancillary feature). This contrasts with the definition of “large” service, which is based on user counts alone. We also note that measuring the number of users on our services is complex, due to difficulties in defining “user count” and due to the different use and functionalities of Google’s services, and this is likely to lead to significant overcounting (and therefore many services may erroneously be identified as ‘large’). [✁] There is therefore no one methodology for measuring user counts and it depends on the Google service in question, as a user might be a content creator or a user accessing content (but not necessarily both). Furthermore, users can choose to access many of our services when they are signed into an account or when they are signed out. Given our systems and privacy policies, we cannot comprehensively deduplicate within these counts or between them which results in significant overcounting. When calculating average user numbers over a period of time, there are also issues with data retention time frames as well as seasonal fluctuations in user counts. We would therefore suggest that there is some flexibility built into user counting, to reflect the different ways in which services operate and ways in which users can be counted in a reliable way, based on the specifics of that service. a. In order to draw reliable conclusions from the data, and for the reasons given above, we would recommend that any user counting should solely focus on signed in users and exclude signed out users. b. We also consider that user counting should focus on users generating content, rather than users accessing content, within the user-to-user part of the service.

Question (Volume 4)	Your response
	c. Given data retention policies (which differ across products), there should be some flexibility over requirements for determining monthly user counts and over what period they should be measured. We note that Ofcom has proposed that users should be counted over a period of 12 months, however, it is rare that our services would hold this volume of data, due to data retention policies, and we would suggest that a six month period would give sufficient evidential basis to determine whether a service meets the threshold for a large service. We also note that a six month period would align with DSA requirements,[10] so anything exceeding this period would require significant changes to systems and processes that have been developed for DSA compliance. d. Given the complexities involved in collecting user counts, the extensive resources required to complete user counting across all services, and the regulatory burden for Ofcom in collating and reviewing the data, it would be proportionate for user counting to be provided as a one-off when a service reaches the relevant user count threshold or for a service to 'self-certify' that it meets the relevant threshold without needing to provide user counts, rather than ongoing regular user counting reporting for all in-scope services. e. Lastly, given the range of users and services, clear guidance of which services are required to proactively provide user counting would be helpful. Suggested amendment We would suggest greater flexibility in both the U2U and Search Codes of Practice as to who is to be considered a user for the purpose of calculating whether a service is large. In particular, we recommend the following guidance is added to the section in both Codes that deals with calculation of user numbers (currently

[10] Article 24(2)

Question (Volume 4)	Your response
	paragraphs A11.7-A11.11 of the draft U2U services Code and paragraphs A8.6.-A8.10 of the search services code): a. a higher user count threshold, given the likely significant overcounting for monthly users; b. building in 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 illegal harms, for example, based on being a signed in user; c. a shorter period for assessing monthly users (e.g. 6 months rather than 12); and d. an ability for services to 'self-certify' that they meet the threshold for a large service, without having to provide precise user counts.
Question 15 (11.4): Do you agree with our definition of multi-risk services?	**Confidential: N** The current definition of "multi-risk" is very broad, as it covers any service that is at least medium risk in relation to at least 2 kinds of priority illegal offences. In our view, it should be limited to those designated as high risk in relation to those offences. Without this change, there would be insufficient delineation between the treatment of services that are medium risk for an illegal harm and services that are high risk for that harm. This could also disincentivise platforms to reduce risk from high to medium as the compliance burden would be the same or very similar. Proposed amendment: As set out at Question 13, we suggest that the definition of a "multi-risk service" be altered to focus on a) residual risk, as described above; and b) services assessed as posing one or more high "specific risk" of harm.
Question 16 (11.6): Do you have any comments on the draft Codes of	**Confidential: N** We made some comments in the opening remarks to our response but look to provide more detail below. **Flexibility**

Question (Volume 4)	Your response
Practice themselves?[11]	One of our concerns is that the Codes are overly prescriptive, and do not build in enough flexibility, depending on type of service or type of harm (e.g. fuzzy keyword detection). In our view, the Codes should permit more flexibility for platforms to design product solutions that best address the issue outline, which would optimise existing expertise and investment. The prescriptive requirements also fail to recognise differences between services: for example, some aspects only translate well to social media platforms, rather than VSPs, due to the ways in which users interact with each other on the service. In general, it would be preferable for the Codes to outline the aim (e.g. minimise the appearance of CSAM on the platform) and provide illustrative examples (rather than explicit requirements) for how this could be done. Failing to do this would result in three extremely negative outcomes: 1) Platforms choosing between two potential safety measures aimed at addressing a certain harm (e.g fraud) may be incentivised to implement the less effective measure on the basis that it matches the solution specified in the Code to demonstrate compliance. Longer term, this would disincentivise services from going beyond the specific proposals in the Code; 2) Platforms that choose to implement alternate, *more effective* safety measures, may find themselves penalised through being unable to benefit from the legal safe harbour; 3) It 'freezes' compliance solutions at this specific moment in time, rather than providing principle-based rules that are future-proofed, and can flex to accommodate new and more effective measures to tackle harms. In the event Ofcom cannot make changes to the face of the Codes to address this issue, and services are required to rely on **alternative measures** for compliance, Ofcom should ensure that a clear mechanism is in place for platforms to implement/record alternative measures and that the process for evidencing alternative measures is not unnecessarily burdensome on the services. We provide further illustrations in our response to **Q18.**

[11] See Annexes 7 and 8.

Question (Volume 4)	Your response
	Proportionality Proportionality should be further baked into the Codes, to ensure that services retain some discretion to apply measures where it is proportionate to do so, for example based on the output of their risk assessments, or based on the functionality of the service. For example, if a platform is considering building a classifier to detect CSAM, it may be more effective to focus detection on forms of content where the greatest risks arise (e.g. video), rather than looking to build a classifier that also covered lower risk areas (e.g. text). This proportionality framing is also set out in the OSA itself.[12] Suggested amendments: ● We recommend that, especially where Ofcom is suggesting more prescriptive compliance measures, each time it should expressly include a proportionality qualification to avoid platforms diverting resources from high/medium risk concerns to low-risk concerns.
Question 17 (11.7): Do you have any comments on the costs assumptions set out in Annex 14, which we used for calculating the costs of various measures?	**Confidential: N** The costs of complying with a specific recommendation will vary greatly depending on a number of variables (some of which are set out below). We cannot therefore comment meaningfully on the cost assumptions set out in Annex 14, except to note that the approach Ofcom has taken is excessively broad-brush and generalised. We do not think it is possible to generate reliable cost assumptions in this way, as opposed to making cost estimates on a more individual basis, factoring in the nature of the relevant service provider, its processes, and its resources. We expect that the costs of a particular measure will vary significantly due to a number of factors, including: ● Location of employees: average wage rates differ significantly between, for example, the US and India. This means that the estimates used in the "Labour Costs" section of Annex 14 are, for certain locations, likely to be materially different to the true costs; ● The complexity of the compliance measure – for example, the "ongoing annual maintenance cost" associated with a

[12] e.g. where services are required to use proportionate measures to target illegal content (s10(2), (3), (4)).

Question (Volume 4)	Your response
	change to systems to comply with a measure is assumed by Ofcom to be 25% of the initial costs but, while this may be accurate for certain measures, the actual cost will vary greatly depending on, for example, the engineering or trust and safety personnel resource required to maintain a compliance system; ● How sophisticated existing systems are: if significant changes are required to update existing processes due to their complexity or scale, associated costs will clearly be much higher; ● Level of resource within the product: there is a huge disparity between products in terms of engineering resource, and overall number of employees (for example, between a product like Kaggle, versus a service like YouTube). In respect of smaller services, any changes made will therefore be likely to take longer and cost more due to the relative lack of resources. A failure to take these factors into account in a more specific way has the effect of making cost estimates inaccurate and unreliable in all but a limited number of circumstances. We would therefore welcome Ofcom recognising in its approach to regulatory oversight and enforcement generally, that many factors will be relevant to the proportionality of what is expected from service providers when complying with the measures recommended in the Codes of Practice.
Content moderation (U2U)	
Question 18 (12.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: N** As we have said in our opening remarks, and at various parts of our response, it is our view that the draft 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. Some areas where we believe the Codes need to be updated to reflect this include: **General monitoring**

Question (Volume 4)	Your response
	We would encourage Ofcom to explicitly set out that none of the duties in the draft code would oblige platforms to generally monitor the service for illegal content (e.g. fraud). This was certainly not the intention during the legislative phase, as confirmed from the despatch box, and we understand that is not the policy intent either. However, we believe further clarity from Ofcom would add regulatory certainty and reduce the risk of significant over removal of legal content. Suggested amendment: We would suggest updating the text to ensure that the prohibition on general monitoring is carried by explicit reference through the Codes, confirming the policy intent. In particular, for each automated content moderation requirement there could be a statement that: *"For the avoidance of doubt, the requirements do not amount to a requirement to conduct general monitoring."* Further, requirements that could be interpreted as general monitoring should be removed or reframed (e.g. requirements to "*analyse all relevant content present on the service at the time the technology is implemented…*")[13] **Content Moderation Policies** We recommend removing the reference to 'emerging harms' in A4.9, Annex 7. Analysis of emerging harms should be part of the risk assessment, rather than a standalone factor in policy development. To provide more context, YouTube invests significant resources to ensure that we are tracking emerging trends on online platforms, not just YouTube. Our Intelligence Desk monitors the news, social media and user reports to detect new trends surrounding inappropriate content, and works to make sure that our teams are prepared to address them before they can become a larger issue. However, we believe this consideration is more appropriately covered in risk assessments. It may be that there is harmful, trending content observed on non-Google platforms. Whilst it is appropriate to monitor these developments and consider preparedness, practically policies cannot consider the nuance of how these harms may manifest until we can assess how it may impact our own platforms.

[13] See A4.25(a), Annex 7; A4.39(a), Annex 7; A4.47(a), Annex 7.

Question (Volume 4)	Your response
	Performance Targets A4.11 (Annex 7) requires services to set and record performance targets covering at least the time that illegal content 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 and illegal under UK law, thus hindering existing processes and mechanisms to evaluate the efficacy of existing content moderation measures. YouTube, for example, measures illegal content through published "Violative View Rate" (VVR) in respect of YouTube content. The VVR shows how many times content has been viewed before it is removed for breaching our policies. We see these VVR as our "North Star" for measuring our progress in combating harmful content and, although not perfect as a metric, is significantly more meaningful than a simple 'time on platform'. We believe that sharing these VVR with the public is an important way to create accountability. 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. Suggested amendment: We recommend that Ofcom: ● We would recommend revising the language of the code (A4.11, Annex 7) to allow platforms to define their own metrics, as long as they can provide reasoning for how this is an effective measure (for example, YouTube measures effectiveness through examining the violative view rate and the % of appeals that are reinstated as a measure of accuracy). ● To the extent that turn around times are referenced in the Code, it would be helpful to clarify that the measures recommended in A4.11 and A4.12 are based on the time from the report of the illegal content to the action taken in response. Currently, the code could be interpreted to mean that the content includes all illegal content, even if it

Question (Volume 4)	Your response
	has not been reported, which implies that service providers need to undertake general monitoring. **Prioritisation framework** A4.15, Annex 7, requires 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. For example, CSAM and fraud might have different assessments, to reflect the factors that are present when identifying this type of content; and ● the type of product, to reflect the differing functionality and risk profiles. For example, YouTube might have a different framework to Google Docs. Suggested amendment: We would suggest that the requirement should be to ensure that the service has an appropriate prioritisation framework that can be explained and evaluated by reference to the risks specific to the relevant service, rather than prescriptively setting out what the framework should be. Examples of aspects that might be included in an appropriate framework could include matters such as the severity of the content, but having regard to these in the framework shouldn't be mandated specifically as they may not be relevant. There are therefore two amendments Ofcom could make at A4.15. 1. The requirement could state that "In setting the policy, the provider **may (where relevant to the service) have regard to**:"; or 2. Ofcom could replace (a) - (c) and instead say that "In setting the policy, the provider **should have regard to appropriate metrics which assess the potential severity of harm being caused by the content"**.
Search moderation (search services)	

Question (Volume 4)	Your response
Question 19 (13.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: N** We consider there are certain areas where the Codes should be updated to reflect how services currently function and better meet Ofcom's policy expectations, as set out below: **Deindexing vs. delisting** The Code refers to search services "deindexing" content. However, we generally "delist" content from the index, rather than "deindex". Deindexing (i.e., deleting content from the search index) doesn't enable the kinds of flexibility that our products rely on, and which is also demanded by other obligations in the Codes, e.g. in the event of a successful appeal, if content has been blocked from serving through delisting, it can be reinstated to results immediately. Since Google operates one index for all of its country services, with those services providing locally-relevant results drawn from that index, a deletion from the index would have the effect of a global takedown imposed as a result of one country's law. This does not appear to be the policy intent of the OSA. If the situation were reversed, UK users could find that UK-lawful content has been suppressed for them based on the laws of other countries that do not share the UK's legislative principles. Proposed amendment: ● We note that the definition of "Deindexing" in the Glossary (at Annex 16, page 80), defines the term "*Deindexing (or delisting)*". As above, we consider these terms to refer to different actions. The Glossary suggests that the use of the term "deindexing" in the Code may be intended to refer to both deindexing **and** delisting as described above and thereby give search services the flexibility to decide whether to deindex or delist. If this is the case, we would recommend clarifying this in the Code by noting that the search service has the flexibility to decide whether to deindex or delist, where required to do so by the Code. Deindexing should only be prescribed for actual child sexual abuse material. Codes should replace deindexing with delisting in all other places. ● We would also suggest amending the definition of "*Deindexing (or delisting)*" in the Glossary as follows: "*Action taken by a search service, where relevant, which*

Question (Volume 4)	Your response
	involves removing a URL from a search index ***or from lists of displayed search results*** *such that it can no longer be presented to users in search results.”* **Delisting vs. downranking** It is not clear to us what “downrank” means under the draft Codes, 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 a 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? 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. The Code requires services to take into account the “prevalence of the illegal content” before taking a decision as to whether to delist or downrank. However, search services are not able to determine “prevalence” of content, as they don’t host the site and do not record metrics like violative view rates. We also note that the requirement for search engines to either deindex or downrank seems to depend on whether the webpage contains “*only a small amount of less severe illegal content and a large volume of valuable lawful content*” (Vol 5, p60). We would recommend that services are not required to downrank in circumstances where a URL contains any amount of unlawful content, and instead give services the discretion to delist. Further, we would want Ofcom’s Codes to be explicit that any removal request is at the URL level to avoid the risk of over removal. Domain-based actions should be limited to “downranking” or demotions. We apply demotions to domains with a disproportionate density of violative material. For example, sites with a high rate of NCEI reports will receive a penalty demotion, over and above the removal and reporting actions we take for individual URLs.[14]

[14] Please see: https://developers.google.com/search/docs/appearance/ranking-systems-guide#removals

Question (Volume 4)	Your response
	Suggested amendment: ● In A4.4, providers should be given the choice as to whether to delist or downrank without having to assess the listed factors (for example, they should be able to delist in all cases, where appropriate). We would suggest that the factors that service providers should have regard to when deciding whether to delist or downrank should not be prescriptively set out. Examples of factors that may be considered could include the prevalence of illegal content and the interests of users, but this should not be mandated. A4.4 could therefore state that: ○ "In considering whether to deindex or downrank the search content concerned, the provider **may (where relevant based on the nature of the content)** have regard to:"; or ○ Ofcom could replace (a) - (c) and instead say that "In considering whether to deindex 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 lawful material that would be affected". ● We also 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. **Prioritisation framework** As above, while we agree that services should have an appropriate prioritisation framework for reviewing removal requests or complaints, we do not believe that certain factors should be mandated by Ofcom in the Codes (such as how frequently search requests are made) as it is not always necessary or appropriate to consider these factors, for every type of harm or every type of search service. Such a prescriptive approach could ultimately slow down the review process. Suggested amendment: We would suggest that the requirement should be to ensure that the service has an appropriate prioritisation framework that can

Question (Volume 4)	Your response
	be explained and evaluated, rather than prescriptively setting out what the framework should be. Examples of aspects that might be included in an appropriate framework could include matters such as how frequently search requests for the relevant search content are made, and severity of the content, but these shouldn't be mandated specifically. We suggest that A4.16 is amended to provide: 1. "In setting the policy, the provider **may (where relevant)** have regard to:"; or 2. Ofcom could replace (a) - (c) and instead say that "In setting the policy, the provider **should have regard to appropriate metrics which assess the severity of harm being caused by the content".** **Performance targets** At A4.12, Annex 8, the draft code requires providers to set and record performance targets for its search moderation function, covering the time that illegal content remains on the service before it is deindexed or downranked; and the accuracy of its decision making. For Search, this obligation 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. More specifically, there is a lack of clarity in the draft Codes about how to measure performance targets. We assume that the time starts to run once the service receives a valid removal request, but this could be clarified. Moreover, it is unclear whether content is "on the service" if it's theoretically eligible to be served but never is. Suggested amendment ● Instead of referencing "deindexing" and "downranking" specifically in 4.12(a) and 4.13, Annex 8, it would be preferable to refer to "enforcement action", as this more closely ties the action taken by the service with the illegality of the content, and gives services broader discretion over whether to deindex, delist, or apply some other penalty. ● A4.12(a) should refer to the time that illegal content remains on the service**, once it has been reported,** before enforcement action is taken.

Question (Volume 4)	Your response
Automated content moderation (U2U)	
Question 20 (14.1): Do you agree with our proposals? Do you have any views on our three proposals, i.e. CSAM hash matching, CSAM URL detection and fraud keyword detection? Please provide the underlying arguments and evidence that support your views.	**Confidential: Y** [⍰] We have set out above our concerns about the more prescriptive elements of the Codes, and these are very applicable to the three proposals below. To emphasise, we agree that the three proposals below are often **one potential compliance solution** to the Act requirements, and the underlying harm. However we have major concerns that these are presented as effectively the only technological compliance solution which will result in platforms falling within the legal safe harbour. In particular we believe Ofcom should avoid proposals which could: i) Disincentivise solutions that go 'above and beyond' the specific proposals; ii) Penalise platforms that implement alternate, more effective solutions; iii) "Freeze' compliance solutions in time - rather than providing principle-based rules that are inherently future-proofed. We have been more specific on these issues below. **CSAM hash matching and URL detection** We recognise that the use of hash-matching to combat CSAM is currently industry standard and indeed it forms a core part of our risk mitigation strategy for many Google services, including YouTube. However, we are concerned that mandating specific forms of technology in the Codes is not future-proofed. It could result in the Codes becoming outdated and may even de-incentivise platforms from innovating and updating more advanced tools to combat CSAM. For instance, if in the future a more accurate technology is developed to detect CSAM, there may be a perverse incentive for companies not to adopt this feature or delay it until either Ofcom updates the Codes or companies receive the necessary assurance from Ofcom that it would meet their requirements On a separate note, we note that the CSAM requirement states that the technology must "*analyse all relevant content present on the service at the time the technology is implemented within a reasonable time*" (A4.25(a), Annex 7). This also applies to A4.39(a),

Question (Volume 4)	Your response
	Annex 7 (the requirement relating to CSAM URLs) and the detection of fraud (A4.47(a), Annex 7) This suggests general monitoring is required, which we understand is not Ofcom's intention. We therefore suggest it should be clarified to explain general monitoring is not required and/or limited to analysing content at certain trigger points (e.g. sharing, generating), and removing subsection (a) of the above provisions. Also see our response to **Q27** below. <u>Proposed amendments:</u> ● We recommend an express statement in the Code that these measures are only illustrative examples of how the safety objective can be achieved - e.g. include wording to state that "platforms should implement the following safety measures, or alternative proactive content moderation to combat CSAM which are at least equally effective" ● Require platforms to apply hash-matching and URL detection tools upon upload of content but not include a requirement to proactively monitor content across the entire service on an ongoing basis. ● Add language in the code that the requirements do not amount to/require general monitoring; and/or remove subsection (a) from A4.25, A4.39, and A4.47 of Annex 7. **Keyword detection for fraud** The draft Code requires certain services to use fuzzy keyword detection to detect fraudulent content. For the reasons set out above, this shouldn't be the only acceptable method of compliance that results in a safe harbour for platforms. Requiring platforms to rely on fuzzy keyword detection risks taking resources away from other, more effective mechanisms for content moderation. Keyword detection can be circumventable and is not always sustainable. Across our products, we use alternate measures, leveraging our many years of continued investment in machine learning technology and refined with years of experience and data input, to meet the policy objective including: anti-fraud protections and classifiers that look for suspicious patterns that make it harder to circumvent than using keywords. [⍰] The Codes should not create a perverse incentive for firms not to invest in safety technology and continue raising the bar for online safety. As currently drafted, there may be a possibility that platforms choose to do the bare minimum to enjoy the 'safe

Question (Volume 4)	Your response
	harbour' provided by the Codes rather than taking the risk of pursuing an alternate approach - even if they can demonstrate that it is more robust. Suggested amendment: We recommend an express statement in the Code that these measures are only illustrative examples of how the safety objective can be achieved - e.g. include wording to state that "platforms should implement the following safety measures, or alternative proactive content moderation to combat fraud which are at least equally effective"
Question 21 (14.2): Do you have any comments on the draft guidance set out in Annex 9 regarding whether content is communicated 'publicly' or 'privately'?	**Confidential: N** We welcome the provision by Ofcom of guidance that seeks to help service providers understand whether content on their service is communicated 'publicly' or 'privately' when taking measures set out by Ofcom in the Code that apply only to content communicated 'publicly'. We also recognise that, in the context of Ofcom's powers to include in the Code measures requiring the use of 'proactive technology' (measures which cannot recommend the use of technology which analyses user generated content communicated privately), Ofcom must have particular regard to the three factors set out in section 232(2) of the Act. In relation to the first and second factors (the number of individuals able to access the content and restrictions on who may access it by means of the service), we think the guidance takes an overly broad and unclear approach in stating "*The fact that there are access restrictions on a service does not necessarily, by itself, mean that content on that service is communicated 'privately'. Ofcom would still expect a service provider to consider how many individuals in the UK are able to access the content...*". We think clear access restrictions, which mean content communicated on a service is intended to be shared with a limited group of individuals, should always mean the relevant content is communicated 'privately'. On a number of our services, users apply access restrictions precisely because they regard that content as private and want it to be shared with a restricted group. There should therefore be no question about whether these communications are public or private. The approach taken

Question (Volume 4)	Your response
	by the guidance would mean that privately stored files would be at risk of automated scanning (due to the application of the automated content moderation measures listed in the Code at A4.22, A4.36 and A4.44). This would have significant adverse implications on users' expectations of privacy and their freedom of expression, given many would be concerned about the use of monitoring tools in relation to their private affairs.
Question 22-24 (14.3): Do you have any relevant evidence on: - The accuracy of perceptual hash matching and the costs of applying CSAM hash matching to smaller services; - The ability of services in scope of the CSAM hash matching measure to access hash databases/services, with respect to access criteria or requirements set by database and/or hash matching service	**Confidential Y** [✂] [✂] Our systems perform well because, at Google, we do not apply each technology in isolation. Rather, the technology is usually combined with other technologies and techniques to ensure accuracy and quality results. We have fine tuned the system and matching thresholds to limit false positives and have built additional safeguards that reduce errors, which may include [✂] [✂]NCMEC specifically hosts hash-sharing databases where hashes are contributed by industry members and specialist NGOs from around the world. These repositories serve as a starting point – but we also [✂] - [✂] - Interested organisations can apply to become CSAI Match users through a form on our external website. ○ [✂] ○ [✂]. - [✂] ○ From a technical standpoint, there are multiple things to consider when adopting and maintaining hash matching capabilities. [✂]

Question (Volume 4)	Your response
providers; and - The costs of applying our CSAM URL detection measure to smaller services, and the effectiveness of fuzzy matching for CSAM URL detection	
Question 25: The costs of applying our articles for use in frauds (standard keyword detection) measure, including for smaller services; and	**Confidential Y** [⍰] In relation to fraud, we are concerned that Ofcom's suggestion to use fuzzy keyword matching would lead platforms to adopt an approach that is less effective than the uses of other technologies they have developed over time to detect this type of content. [⍰] By being prescriptive about the use of specific and prescriptive technologies, Ofcom risks disincentivizing platforms to adopt the most effective technologies for their platforms. This, coupled with the risk of losing a legal safe harbour, could realistically create a race to the bottom, where platforms adopt an approach to demonstrate compliance, rather than continuing to innovate as bad actors adapt. We believe this runs contrary to the objectives of the OSA.

Question (Volume 4)	Your response
Question 26: An effective application of hash matching and/or URL detection for terrorism content, including how such measures could address concerns around ‘context’ and freedom of expression, and any information you have on the costs and efficacy of applying hash matching and URL detection for terrorism content to a range of services.	**Confidential: N** Hash-matching technology is technically possible for terrorist and violent extremist content. It is most reliable where it is used to detect reuploads of violative content (i.e. content that has previously been determined by a platform to be violative of their policies). For example, a platform can use in-house hash matching technology to detect copies of content their human reviewers have already determined to be violative of their own content policies. This helps to scale up a content moderation decision across the much wider corpus of content hosted by a platform. It is also possible to use hash matching to detect known violent extremist content from GIFCT’s hash-sharing database. However, the hashes shared by GIFCT member companies are not necessarily illegal, but rather are violative of individual companies’ voluntary (and varying) content policies. This is due to the fact that there is no international consensus on how to define “terrorist content”. What this means is that hashes can assist platforms in detecting potentially violative content, but there are limitations on the utility of the hash database. For example, a hash match does not necessarily lead to content removal if, upon review, it does not violate a platform’s policies. GIFCT’s hash-sharing taxonomy establishes consensus between member platforms on general and behavioural inclusion parameters, terrorist entities based on the UN Security Council’s Consolidated Sanctions List, and criteria for the Incident Response Framework (IRF). It is also worth noting that GIFCT is a membership based organisation and the hash sharing database is not open to non-members. To become a member, technology-based companies need to apply and demonstrate they meet the membership criteria. Once approved as a member, eligible platforms are then permitted to access the hash-sharing database. GIFCT members are not obligated to utilise the hash sharing database, and not all members participate.
Automated search moderation	

Question (Volume 4)	Your response
Question 27 (15.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: Y** [✁] **CSAM URLs** A4.26(b) is currently phrased so as to require providers to regularly monitor CSAM URL lists and "identify CSAM URLs that have been removed from the list and reinsert them into the search index". This obligation to monitor lists and to reinstate CSAM URLs removed from authoritative external lists is problematic, as this amounts to a "must carry" provision. Service providers should have agency to determine reinstatements, particularly in cases where violative material (for CSAM or other legal/policy reasons) remains on the page. Moreover, since we deindex, rather than delist, CSAM, it would be technically overly burdensome to reinstate the content item in question. Suggested amendment: We suggest that Ofcom's Codes should include a provision that clarifies that platforms have agency to not re-insert content in the search index if it violates their content guidelines or Terms of Service. **CSAM warnings** Our current Search CSAM warnings are not triggered in the UK by a static list of CSAM keywords, but are triggered by a combination of classifiers and query understanding. ● [✁] Also, as currently drafted, the code requires services to include links in the warnings to resources designed to help users refrain from committing CSEA offences that are freely available through a reputable organisation dedicated to tackling child sexual abuse. This seems overly prescriptive, giving little room for innovation/adaptation. Suggested amendment: The Codes should empower service providers to find the most effective way to deploy CSAM deterrence messages, using triggers appropriate to the service. Also, instead of referring to "reputable organisation dedicated to

Question (Volume 4)	Your response
	tackling child sexual abuse”, an alternative reference to “support resources” would provide adequate flexibility.
User reporting and complaints (U2U and search)	
Question 28 (16.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: Y** [⍰] **<u>Complaints and Appeals (Search) - Annex 8</u>** **Reversing decisions on appeal** The draft Code states that if a service reverses a decision on appeal, the search content should be restored to its previous position. 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 illegal, the provider is asked to “restore the search content to the position it would have been in" had it not been judged to be illegal content. 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 be illegal, it might be policy violative and the service provider should still retain the ability to demote or delist content if it is harmful or of low quality, if it violates content policies. 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 to prevent similar issues. However, we feel 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. <u>Suggested amendment:</u>

Question (Volume 4)	Your response
	- Annex 8 should reflect the principle that content should not be penalised whilst under review. - An amendment should be made to A5.18(a), so that the provider should "*use reasonable endeavours*" to restore the search content. - Annex 8 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. **Complaints system (Search) - Annex 8** On the provision regarding the number of clicks 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. <u>Suggested amendment:</u> - We recommend amending A5.4(c), Annex 8 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 demotions (Search) - Annex 8** We consider this provision to be problematic. As currently drafted, the draft Codes 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. In particular, much of the way Google Search limits the risk of

Question (Volume 4)	Your response
	problematic content appearing in search results, especially for queries that are not explicitly seeking it, is closely tied into Search's core mechanisms for assessing the overall quality of content, a concept that encompasses but goes well beyond the kinds of content risk at issue here. If that leads to a conclusion that Search's core quality signals are "proactive technology" whose application webmasters can appeal, that will create a serious structural issue for Google. Search appeals should therefore be limited to delisting, rather than ranking. It should also be limited to illegal content, which, in line with our policies, will be delisted rather than demoted. [⍰] In other jurisdictions, we have been actively prohibited from notifying webmasters of delistings applying to their content in Search, for instance on the grounds that the underlying request (or even the affected pages) contain the personal data of the requester. It's very important that any appeal process, including the notification to a site owner telling them of a moderation action, be able to disclose a reasonable amount of information about the basis for the action, which will often include information about who complained and on what basis. If, on the other hand, it's determined that such information should not be provided to protect the requester, then the service provider should be explicitly relieved of the obligation to notify and handle appeals. Suggested amendment ● Clarification in the Code that service providers can tie complaints processes to the compliance measure adopted by the service. This means that where service providers downrank illegal content, the only complaints process available should be relating to ranking t; and when a service provider delists illegal content, the only complaints processes available is that relating to delisting. ● The Code could also specify that where a complaint is clearly frivolous or vexatious on its face, the service provider does not need to take any action.

Question (Volume 4)	Your response
	Appeals (U2U) - Annex 7 As noted in our Exec Summary, the Codes should clarify that the proposed measures should be implemented in a proportionate manner in order to achieve the best safety outcome for users. [✂] In this way, a users' investment in commenting on a video does not compare to a creator's investment in terms of time and resources in creating the video content available on the VSP, and there is no expectation that the comment will remain live indefinitely. As such, there is far [✂] [✂] **Content Restoration - Annex 7** The Codes require restoration of user content to the position it would have been in had the content not been judged to be illegal (see Vol 4, 16.136 p. 194; Annex. 7, A5.18). There may be situations where content restoration is not possible (e.g. because the content is ephemeral/time-sensitive by nature or technical limitations in the design of the platform mean that it is not reasonably feasible to "restore" content). Suggested amendment Amend the Codes to clarify that restoration is not required where restoration is not reasonably feasible or, content is by its nature ephemeral/time-sensitive. **Both Search & U2U (Annexes 7 and 8)** **Trusted flaggers** ● In order for the system to work effectively and expeditiously, the Code should require trusted flaggers to include details of why the content is illegal, and not just report the content alone, in order to distinguish the process from a user flag. Further, platforms should be able to assume that where Government, regulators or other trusted flaggers report allegedly illegal content, that they have carried out basic publicly-available checks e.g.

Question (Volume 4)	Your response
	FCA-authorised entities, and can provide such evidence to the service. **Supporting material** The Codes can be read as contemplating that all complaints mechanisms should enable users to provide supporting material (see Vol 4, p. 171; Annex. 7 & 8, A5.4). There may be situations where it is not necessary or proportionate to provide users with the ability to submit supporting material in addition to relevant text-based information, noting additional technical burdens associated with building such functionality for all complaints systems. For example, many Google products enforce policies regarding “obscene and profane content”-- a requirement to enable users to provide supporting documents in addition to relevant text-based information when complaining about the removal of such content is likely to be disproportionate in many circumstances given a user’s ability to adequately state their complaint without additional supporting material. Suggested amendment Amend the Codes to clarify that users and affected persons should have the ability when making relevant complaints to provide the provider with relevant information or supporting material, but only to the extent reasonably appropriate in the circumstances. **Prioritisation framework** We note that for large or multi-risk services, the provider should prioritise appeals, based on the (i) the severity of the action taken against the user; (ii) whether proactive technology was used to make the determination; and (iii) the service’s past error rate in relation to illegal content judgments of the type concerned. Suggested amendment: 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 (see further **Q18 and 19)**.

Question (Volume 4)	Your response
	Avoiding similar errors The Codes (see Annex. 7 & 8, A5.18) use the language “where necessary to avoid similar errors in future” as a condition to certain obligations. First, it is not necessary to include that language as a condition to paragraph A5.18(c) which reads more clearly without it (because the point is covered by the proviso that the technology “does not cause the same content to be taken down again”). Second, the requirement to “where necessary to avoid similar errors in future, adjust the relevant content moderation guidance” does not take account of the fact that it is not possible to avoid all errors in content moderation at scale, despite the best efforts of providers such as Google (see Volume 4, [12.2]], where Ofcom recognises: “*We know that content moderation systems, particularly those deployed across a very large user base, cannot provide a guarantee that users will not encounter any illegal content (and are often designed around reducing instances, rather than complete prevention)*.” The systematic improvement of content moderation at scale is dependent on the ability of providers to make system and process adjustment in response to trends in data rather than individual instances of error. Suggested amendment Remove the words “where necessary to avoid similar errors in future” from A5.18(c) in Annexes. 7 & 8. Remove the requirement to adjust relevant content moderation guidance in response to individual decisions and instead include a general obligation to review content moderation guidance periodically having regard to the outcome of appeals. **Spam and providers of malware** We are committed to dealing with all genuine complaints in accordance with the Act. 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.

Question (Volume 4)	Your response
	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 Codes do not presently recognise that "*appropriate action*" to take in respect of spam and complaints by providers of malware should be materially different to genuine complaints. For example, a requirement to acknowledge spam complaints (see [A5.9]) would impose a significant burden on service resources. Similar issues arise throughout the balance of the Reporting and complaints guidance. Suggested amendments Amend the Codes to clarify that relevant substantive obligations outlined in the Code do not apply to spam or complaints by providers of malware. **Drafting clarification** Section 5H in Annexures 7 and 8 contain a requirement that services deal with "all other relevant complaints" in accordance with Recommendations 5D - 5G. However, Recommendations 5D - 5G are distinct types of complaints. The plain language of 5H reads as though in the event 5H applies services should comply with all recommendations in 5D - 5G, rather than just the most relevant category of recommendations. Based on the explanation of this Recommendation in Volume 4 at [16.142]-[16.144], we understand that Ofcom intends for providers to follow the most relevant category of recommendation only. Suggested amendment Amend the Codes to clarify that for "all other relevant complaints" services should deal with the complaint in accordance with the most relevant category of recommendations in 5D - 5G.
Terms of service and publicly available statements	

Question (Volume 4)	Your response
Question 29 (17.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: N** We recommend Ofcom includes an express recognition that platforms should only be required to provide a level of detail regarding their automated technology (or other measures that address priority illegal content) that does not jeopardise the effectiveness of those measures.
Question 30 (17.2): Do you have any evidence, in particular on the use of prompts, to guide further work in this area?	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.
Default settings and user support for child users (U2U)	
Question 31 (18.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: N** We support the intention to provide support to child users so they are protected from encountering illegal harms such as CSAM and grooming. It is worth noting that in a number of places the draft Codes seem to incorrectly reference YouTube as a social media network, whereas in fact it is a Video Streaming Platform and - like many other U2U services that will be in scope - has different functionalities than social media and would therefore have a different level of risk of harm. For instance, YouTube does not have a direct messaging function nor network expansion prompts.
Question 32 (18.2): Are there functionalities outside of the ones listed in our proposals, that should explicitly inform users around	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.

Question (Volume 4)	Your response
changing default settings?	
Question 33 (18.3): Are there other points within the user journey where under 18s should be informed of the risk of illegal content?	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.
Recommender system testing	
Question 34 (19.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: N** As we set out in our opening remarks, we are concerned that the framing for how recommendation systems are considered throughout the Codes over indexes on the potential for harm. The code and associated requirements should reflect the ways in which recommender systems can help compliment other protections platforms have in place, like Community Guidelines. For example, recommendations can help connect viewers to high-quality information and minimise the chances they'll see problematic content - this is critical from a platform safety perspective and paramount to our goal of recommending content that delivers value. Additionally, we believe the testing requirements are overly prescriptive and it is unclear how these requirements would actually mitigate risk of illegal content. We note that the Act requires platforms to assess risk prior to any 'any significant change to any aspect of a service's design or operation'[15], and Ofcom has provided further guidance on this. This requirement would impact changes to recommender

[15] s9(4)

Question (Volume 4)	Your response
	systems in the same way as other platform functionalities. It is therefore not in line with Act, and not beneficial from a user safety perspective, for the Code to introduce quasi-risk assessments for recommender system design changes, regardless of how "small and incremental" they may be. The definition of 'significant change' already builds in assessment of end-user risk, so there is no safety-reason, nor any justification under the Act, for Ofcom to lower the 'significant change' threshold for recommender systems. The current Codes also introduce an enhanced compliance burden for on-platform testing. This could create an unintended consequence where platforms are incentivised to only carry out off-platform testing prior to launch. In most cases, this would not be as robust as on-platform testing and may not meet Ofcom's policy intent of reducing the risk of online harms. Suggested amendment: As currently drafted, all product changes to recommender systems are potentially in scope of this obligation. The code should align recommender changes with other system design changes, and require assessments for 'significant changes' to recommender systems. Also, the test for 'significant changes' to recommender systems should allow platforms flexibility to determine the appropriate testing measures and documentation processes that capture any risk associated with illegal content. Ofcom may still think it necessary to suggest illustrative examples, but these should be clearly explained to allow platforms to explore alternative processes.
Question 35 (19.2): What evaluation methods might be suitable for smaller services that do not have the capacity to perform on-platform testing?	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.

Question (Volume 4)	Your response
Question 36 (19.3): We are aware of design features and parameters that can be used in recommender system to minimise the distribution of illegal content, e.g. ensuring content/network balance and low/neutral weightings on content labelled as sensitive. Are you aware of any other design parameters and choices that are proven to improve user safety?	**Confidential: N** Our recommender systems help connect viewers to high-quality information and minimise the chances they'll see problematic content. However, where content is judged to be illegal, we do not allow it on our services and we remove it.
Enhanced user control (U2U)	
Question 37 (20.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: Y** [✂] **General Comments** At a high-level, we have serious concerns with these proposals, at least at this stage and form. In particular: (i) the provisions are too prescriptive and, for the reasons detailed in the Executive Summary above, may not be the best means for individual platforms to meet the underlying harms; (ii) The Act specifically provides details on what types of user controls platforms should provide, and which platforms should be

Question (Volume 4)	Your response
	obliged to do so, through the 'User Empowerment duties'[16]. Parliament could have chosen to include these 'Enhanced User Controls' in its 'User Empowerment' duties, which specifically covers the obligation to allow users to 'block' anonymous users, but chose not to. By adding these obligations within the Codes, Ofcom is effectively going beyond the scope of the Act; (iii) Despite these reservations above, to the extent Ofcom does include 'Enhanced User Controls', it should do so alongside the User Empowerment Codes in Phase 3, given the clear overlap with those obligations. **Specific Points / Recommended Changes** We note that Ofcom cites "i) grooming; ii) encouraging or assisting suicide (or attempted suicide) or serious self-harm; iii) hate; iv) harassment, stalking, threats and abuse; and v) controlling or coercive behaviour" as the key harms which the section aims to combat. We acknowledge that these are important harms for platforms to address, but are concerned that the same provisions apply to platforms with full social-media functionalities (e.g. direct messaging, user connections) as those which have minimal social functionality (e.g. a sports publisher with comments functionality). For the reasons detailed above ('General Points'), we have significant reservations about these provisions being included in the Illegal Content Codes (rather than Codes covering User Empowerment), and we question the substance of the provisions. However, to the extent the provisions remain in some form, we recommend that Ofcom should include clear proportionality provisions, to reflect the fact that not all elements of the Enhanced User Controls are appropriate for all types of U2U platforms. [⍰]

[16] See Section 15 OSA

Question (Volume 4)	Your response
Question 38 (20.2): Do you think the first two proposed measures should include requirements for how these controls are made known to users?	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.
Question 39 (20.3): Do you think there are situations where the labelling of accounts through voluntary verification schemes has particular value or risks?	**Confidential: N** The requirements for notable user schemes and monetized schemes are overly prescriptive and may disincentivize platforms from establishing or maintaining these. For example, ● A9.12.(b)(iii)(iv), Annex 7 - requires a service to establish criteria and thresholds that set out how the provider will satisfy itself that the user account of a relevant user is operated by or on behalf of the person by whom or on whose behalf it is held out as being operated; and if that person is held out as holding a particular position or role, that they hold that position or role. ○ We recommend deleting this requirement does not enable flexibility for platforms who may have different verification/labelling schemes for a wide range of users who hold a wide variety of different positions and/ or roles. It is unclear how platforms would meaningfully meet this requirement to mitigate the intended risks. ● A9.12(c), Annex 7 - requires platforms to *“set out safeguards to ensure that the user profile information (such as username and ‘bio’ text) provided by the relevant user when their user profile was labelled under a notable user scheme is not modified without the provider reviewing and consenting to that change”*. ○ We recommend deleting this requirement. This is far too wide an intrusion on users’ ability to update their own channels/content e.g. For YouTube it would mean that UK users ranging from publishers, musicians or high profile creators, would not be

Question (Volume 4)	Your response
	○ able to update their profiles/bios without prior YouTube authorisation. Take Channel 4 as an example - Channel 4's entertainment channel is verified and in its 'About' profile shares information including some of its most popular shows currently available to watch on the platform; if they decided to change this information, they would need prior YouTube approval. ● A9.12(e), Annex 7 - requires platforms to "*set out how the provider will treat relevant users and the content they post on the service, including recommender systems, content curation, user reporting and complaints, quality assurance, fact checking, content moderation, and account security*". ○ We are generally unclear of Ofcom's intention for including these provisions and we recommend deletion. The Codes and Act already cover most of these issues, and it seems unnecessary to duplicate these provisions (and almost impossible for platforms to implement parallel functionalities in this area if that was the intention).
User access to services (U2U)	
Question 40 (21.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: N** We broadly agree with these proposals. We note our concerns with how the Codes expand on what are 'reasonable grounds to infer' - **see q.49**
Questions 41 -43 (21.2): Do you have any supporting information and	**Confidential: Yes** [✂] [✂]

Question (Volume 4)	Your response
evidence to inform any recommendations we may make on blocking sharers of CSAM content? Specifically: ● What are the options available to block and prevent a user from returning to a service (e.g. blocking by username, email or IP address, or a combination of factors)? What are the advantages and disadvantages of the different options, including any potential impact on other users? ● How long should a user be blocked for sharing known CSAM, and should the period vary depending on the nature	

Question (Volume 4)	Your response
of the offence committed? ● There is a risk that lawful content is erroneously classified as CSAM by automated systems, which may impact on the rights of law-abiding users. What steps can services take to manage this risk? For example, are there alternative options to immediate blocking (such as a strikes system) that might help mitigate some of the risks and impacts on user rights?	
Service design and user support (search)	

Question (Volume 4)	Your response
Question 44 (22.1): Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	**Confidential: Y** [✂] The Code suggests that large search services should offer a means for users to easily report predictive search suggestions that direct to priority illegal content and remove suggestions where there is a risk of users encountering illegal content. [✂]
Cumulative Assessment	
Question 45 (23.1): Do you agree that the overall burden of our measures on low risk small and micro businesses is proportionate?	**Confidential: N** We have significant concerns around Ofcom's apparent assumption that risk is directly proportional to the size of a service without also taking into account other factors. This is most evident in i) the approach to evidence needed for the risk assessments; ii) the approach to assessing impact of harm; iii) the suggested governance measures; and iv) the approach to the significant change trigger for reviewing risk assessments.
Question 46 (23.2): Do you agree that the overall burden is proportionate for those small and micro businesses that find they have significant risks of illegal content and for whom we propose to recommend more measures?	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.

Question (Volume 4)	Your response
Question 47 (23.3): We are applying more measures to large services. Do you agree that the overall burden on large services proportionate?	**Confidential: N** Please see response to **Q45 above.**
Statutory tests	
Question 48 (24.1): Do you agree that Ofcom's proposed recommendations for the Codes are appropriate in the light of the matters to which Ofcom must have regard? If not, why not?	**Confidential: N** We recognise the balance Ofcom has to strike on the specificity of the Codes given the scope and scale of the obligations and the variety of services it will apply to. However, we consider the draft Codes would benefit from practical improvements to ensure Ofcom's policy objectives are met. We note that in preparing the draft Codes of practice and guidance, Ofcom is required by the OSA to consider certain principles, including "*the measures described in the code of practice must be proportionate and technically feasible*"[17]. Furthermore, measures described in a code of practice which are recommended for the purpose of compliance must be designed in light of the following principles: ● The importance of protecting the right of users and (in the case of search services) interested persons to freedom of expression; and ● The importance of protecting the privacy of users.[18] We have raised concerns about the extent to which Ofcom has sufficiently considered these principles elsewhere in our consultation response. In summary: ● We would welcome further clarity from Ofcom on the proportionality of certain obligations (see our responses to earlier questions. Furthermore, where Ofcom mandates

[17] Schedule 4, para 1 and para 2(b).
[18] Schedule 4, para 10(1) and (2).

Question (Volume 4)	Your response
	specific examples, these could be framed more broadly (to allow greater discretion for services), or it should be made clear that services are only required to implement those solutions where it would be proportionate to do so. ● We also consider that greater consideration could be given to the rights of users to freedom of expression, particularly in terms of the guidance relating to illegal content judgments and the likely implications for users where this threshold may lead to over-removal of legal content. ● We have highlighted some concerns about the impact on users' rights of privacy in relation to some aspects of Ofcom's Codes and guidance (see responses to earlier questions), and would like to ensure that the online safety regime is compatible with requirements under data protection laws (such as UK GDPR). In particular, we would welcome clarification in the Codes and guidance that services are not required to collect or process additional personal data from their users in order to comply with the requirements of the Act.

Question (Volume 5)	Your response
Illegal content judgments	
Question 49 (26.1): Do you agree with our proposals, including the detail of the drafting? What are the underlying arguments and evidence that inform your view.	**Confidential: N** We consider the specific illegal content guidance to be very helpful, particularly for smaller services. Our prime concern lies with the guidance around what is 'reasonable grounds to infer' - and the potential of over-removal of lawful content as a result. **Reasonable grounds to infer** Ofcom's guidance envisages that services will need to label content as "illegal" where they have reasonable grounds to infer this, even in circumstances where a court would not do so. We agree with Ofcom's assertion that "there are no criminal implications for the user if their content is judged to be illegal content against this threshold".

Question (Volume 5)	Your response
	Nevertheless, there are other implications for users, such as on the user's rights of free speech, or to monetise content, and therefore, in our view, where platforms are making impactful decisions on 'illegality' of content, the threshold should reflect the seriousness of making such a judgement. In our view it is only 'reasonable to infer' illegality when it is also 'reasonable to infer' that a court would do so. On one interpretation of the existing test, services would be required to remove content unless there is some basis that a user could "successfully" rely on a defence. For example, if a user flags an ad as fraudulent, unless the service can establish that it *isn't*, it will be required to remove it. Ofcom specifically recognises that platforms will be required to over-remove in the context of fraud, and that this could have an impact on UK business's ability to function (see 26.187). It ultimately places a legal obligation on services to remove lawful content, and negatively impacts freedom of expression or ability to monetise content. **Recommendation** ● Ofcom should emphasise that it is only 'reasonable to infer' illegality if it is 'reasonable to infer that a court or equivalent judicial body would judge the content to be illegal'. This recognises the high burden of proof that is generally attached to illegal content judgments under UK law, protects against over-removal of lawful content and thereby ensures freedom of expression is protected to a similar degree online as offline. ● We further suggest that Ofcom clarify in the guidance that such "reasonable grounds to infer" means that a service provider must have a strong evidential basis for concluding that all elements of the offence are made out and that the content amounts to an illegal offence, and, in the event of insufficient certainty or evidence, the presumption should be in favour of the content creator. ● The guidance should also ensure that, in assessing whether something is illegal, services are required to

Question (Volume 5)	Your response
	consider the public interest value in the content (as envisaged by the Act under section 22(2), for example). The guidance should clarify that clear public-interest content should only be removed following a court determination that the content is unlawful.
Question 50 (26.2): Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.
Question 51 (26.3): What do you think of our assessment of what information is reasonably available and relevant to illegal content judgements?	**Confidential: N** **Reasonably available information -** [See A1.66 of Annex 10] Services are expected to make content judgements, based on 'reasonably available information', which the guidance states may include: (a) Content Information; (b) Complaint information (provided with the referral); (c) User profile information; (d) User Profile Activity; (e) Published information (e.g. terror lists) (see 26.26) This is a very broad list of information, and we believe some clear limits need to be placed on this, both to recognise the scale at which services are expected to moderate (since it would not be feasible or proportionate to do this at scale), and due to the potential impact on privacy of these investigative requirements. Even where services do have access to certain information, it may not always be reasonable from a privacy perspective to access User profile information and User profile activity. Services would need to consider carefully whether the processing of this data, to the extent that it constitutes personal data and/or special category personal data, is compliant with the service’s obligations under the UK GDPR and Data Protection Act 2018. In order to

Question (Volume 5)	Your response
	avoid confusion between the application of the UK GDPR and the Act we would welcome this being made clear in the Guidance. In particular: (i) On a data subject basis, it may not always be legitimate/proportionate to access user data in this way*; (ii) At a system process level, it may mean that large platforms would need to have hundreds of staff handling tens of thousands of reports each week. There are inherent privacy/security risks associated with providing user-level data access to all reviewers. * Ofcom seems to recognise this even for fairly extreme cases e.g. in relation to identifying the age of the subject of an indecent image or the victim of a sexual activity offence, noting that “Services should have regard to the privacy implications of reviewing a user’s account activity and information in order to determine their age. This is likely to amount to a very significant interference with their privacy and that of the other users they interact with.”[19] Consistent with Ofcom’s duty under s99(4) OSA, we would also welcome Ofcom consulting the Information Commissioner before finalising guidance on illegal judgement. Suggested amendment ● Revise the Guidance to make it clear that: ○ services are responsible for determining which personal data, if any, they will process as part of their assessment of illegal content; ○ any new collection of personal data would need to be weighed up against the privacy risk and the additional intrusion into privacy be proportionate to the benefits of online safety; and ○ there is no requirement to process special category personal data ● There should be clarity that the information that is “reasonable” for a service to collate and analyse should be

[19] A4.22

Question (Volume 5)	Your response
	proportionate to the nature of the content in question. This includes: ○ the severity/volume of alleged illegal content (i.e, a livestream of a terror event from a purported news outlet requires greater investigation for context than a still CSAM image); ○ the extent to which there is a significant public interest or freedom of expression aspect to the content. ● Services should be able to accept details provided by trusted or government flaggers, without needing to verify this information ● Service should be able to assume that the content reporter has carried out appropriate publicly-available checks e.g. whether the entity is FCA-authorised or not, particularly from a trusted/govt flagger.

Question (Volume 6)	Your response
Information powers	
Question 52 (28.1): Do you have any comments on our proposed approach to information gathering powers under the Act?	**Confidential: N** We are supportive of Ofcom's indication that it will exercise its information-gathering powers in a way that is proportionate to the use to which the information is to be put in the exercise of Ofcom's functions, as required by the Act. While we appreciate that Ofcom will require certain information from regulated services, that must be balanced against the need to ensure that the regulatory burden in responding to information requests is not disproportionate or excessive. The scope and nature of information requests should be targeted and proportionate to the function they are assisting. For example, the broadest, most intrusive and detailed information requests should be reserved

Question (Volume 6)	Your response
	for the most serious enforcement issues, where Ofcom suspects a breach of the service provider's obligations. In contrast, information requests supporting Ofcom's routine regulatory functions should be limited to seeking only the information necessary to carry out those functions effectively. We would welcome this distinction being drawn in the Enforcement Guidance. We also have some specific observations below. **Remote viewing of service in operation / algorithmic testing** In Volume 6 - 28.9, Ofcom refers to its powers to issue information notices so as to remotely view information demonstrating in real time the operation of systems, processes and features used by the service. This is one of the more intrusive information gathering powers available to Ofcom, particularly as access to live user information could interfere with users' rights of privacy and raise security issues where it is accessing commercially sensitive information or information that could be used by bad actors to game our systems. As mentioned above, it is important that Ofcom exercises its powers in a manner that is proportionate. We are concerned that the guidance does not build in appropriate safeguards or provide clarity over when Ofcom will use these kinds of powers. Given the privacy and security implications of this power, in our view it should only be exercised as a matter of last resort, where a service provider is in breach of other information-sharing obligations under the Act, or where the relevant information cannot be obtained by other means. Likewise, Volume 6, 28.22 refers to Ofcom's obligations in relation to ordering a skilled person's report. However, it does not include any criteria for the skilled person (in order to ensure suitability and appropriateness of expertise), any procedure for handling sensitive user data or confidential information, or how to conduct preliminary conflicts clearance before any skilled person is appointed. We would welcome clarity on these issues in the guidance. Proposed amendment

Question (Volume 6)	Your response
	We suggest that paragraph A5.34 clarifies that, in line with Ofcom's proportionate approach to information notices, it will reserve the most intrusive, detailed and wide-ranging information requests to situations where it is investigating a potential breach of the Act's requirements and seeks information for that purpose. In contrast, information notices designed to support Ofcom's other regulatory functions will generally be more limited in nature and scope. We suggest that in Section A5 of Annex 11 on Enforcement wording to the following effect: "*Ofcom recognises the intrusiveness and considerable cost to services of Ofcom's power to issue information notices requiring the real time demonstration or testing of the operation of algorithms, and must exercise its powers proportionately and only where other alternatives have already been exhausted or would not achieve the stated objective*". We suggest that paragraph A5.39 sets out: - Certain criteria for the appointment of a 'skilled person', for example, that the relevant person must have the necessary expertise and demonstrated experience to understand and investigate the service under investigation; - That Ofcom will ensure the person is impartial and will confirm that it has conducted preliminary conflicts clearance, and will set out to the service provider how it has done so, before any skilled person is appointed; and - That the skilled person will confirm that they will handle sensitive user data or confidential information in line with procedure aimed at maximising preservation of confidentiality in the material Furthermore, we recommend Ofcom set out how safeguards will be incorporated into the process, for example by recognising that this is a power that will only be used where Ofcom has first asked for information in a written request and not been satisfied with the response.
Enforcement powers	

Question (Volume 6)	Your response
Question 53 (29.1): Do you have any comments on our draft Online Safety Enforcement Guidance?	**Confidential: N** We understand that in Ofcom's capacity as the independent regulator for online safety in the UK, they will be tasked with ensuring that enforcement is conducted in a way that is fair and proportionate and in line with their public law duties. As such, Ofcom will be guided by their regulatory principles including operating with a bias against intervention, whilst also ensuring that interventions are evidence-based and proportionate. We note that, Ofcom must ensure that in its role of enforcement of online safety, that the least intrusive regulatory mechanism is adopted to achieve these policy objectives. **Corporate structures** We note that the guidance refers to Ofcom holding "*another company within the same corporate group as a service provider liable for a contravention of the service provider's duties under the Act*"(A2.4, Annex 11). We recognise that Ofcom has powers under Schedule 15 of the Act to issue an enforcement decision or notice to both the service provider and related companies. However, we are concerned that the guidance does not reflect the constraints on those powers built into the statute. For example, in respect of subsidiaries of a service provider, the Act makes clear that a relevant decision or notice may only be given to the subsidiary where it "*contributed to the failure in respect of which the decision or notice is given*". Whereas, paragraphs A7.15-A7.20 of the enforcement guidance go further than this and suggest that Ofcom may consider it appropriate to pursue a Related Company, including subsidiaries, not only in situations where the company has some responsibility for the failure under investigation, but also where enforcement action would be more effective if taken against the related company as well as the service provider. In particular, at A7.19 Ofcom states that action may be taken due to "*concerns about the resource required to ensure [a service's] compliance with any confirmation decision that we impose via the mechanisms of another jurisdiction.*" We note that under the Act, and under English law more generally, subsidiaries are not held liable for the actions of parent companies, unless they have

Question (Volume 6)	Your response
	been materially culpable in the infringing conduct, particularly where the basis is primarily due to perceived inefficiencies with enforcing overseas. Clarification amendment To clarify this, we would suggest that A7.18 is also amended to state "*In the case of other Related Companies, we would expect to have some evidence that the other company materially contributed to the failure under investigation…*" **Implementation period** We are grateful that Ofcom is mindful of the likely operational lift required to achieve full compliance with the Act and related Codes, and is allowing services 6 months from publication of the Code of Practice as an implementation period. However, we consider that this may not be sufficient for all services, particularly where the final policy decision has yet to be made or where additional obligations apply depending on size of service and/or conclusions about whether the service is medium or high risk of a particular harm. We would therefore encourage Ofcom to seek a more realistic period of 9-12 months for full compliance; or for services to achieve substantial compliance within 6 months, with a clear roadmap for full compliance thereafter on particular provisions. **Ofcom's priority framework:** Ofcom has provided helpful guidance regarding the circumstances in which decisions will be made about whether or not to open an investigation or take other action against a service provider by reference to certain priority factors. However, A3.9(b)(i), Annex 11 states that as part of Ofcom's priority framework it will consider "*whether enforcement action would help clarify the regulatory or legal framework*". In our view, enforcement should not be used as a means to "clarify" ambiguity in the Codes. On the contrary, where a lack of clarity in the Codes has led to inferior outcomes, we would welcome constructive engagement between Ofcom and the service provider, without the need for formal enforcement proceedings, and that the Code

Question (Volume 6)	Your response
	itself is directly amended. Otherwise services would be unfairly penalised for Codes that lacked clarity. We would therefore ask for A3.9(b)(i) to be removed from the Code. We would also welcome an additional factor being incorporated as part of Ofcom's priority framework (at A3.9). It could be a relevant factor as to whether the service provider has self-reported, or voluntarily notified Ofcom of a safety issue, as part of Ofcom's assessment as to whether to commence enforcement proceedings. Including this as an additional consideration may facilitate a more efficient and constructive resolution of the safety concern, without resulting in enforcement action, as well as acting as an incentive for services to proactively engage with Ofcom where issues are identified. **Ofcom's initial assessment** We welcome Ofcom's indication in the enforcement guidance that it will generally engage with the service provider as part of its initial assessment, to give it an opportunity to comment on the issues. This is an important aspect of procedural fairness, as well as being critical to allowing a service provider to promptly investigate. However, we note at A4.14, Annex 11 that the guidance lists exceptions to this principle, and a broad range of circumstances in which Ofcom may choose not to notify or engage with a service provider when deciding whether to commence an investigation. ● This includes circumstances where Ofcom considers that there is "*sufficient information to conduct [an] initial assessment and decide the appropriate next steps*" (A4.14(a)). In our view, it is difficult to see how Ofcom might have an evidential basis to reach this conclusion, without giving the service provider the right to investigate or respond to the complaint. To do so might give the impression of bias in the investigative process. ● It is also not clear why moving directly to an investigative stage, rather than engaging with a service provider, would be appropriate, simply due to a need to move quickly (A4.14(b)). Conversely, it may help resolve an issue more

Question (Volume 6)	Your response
	quickly if a service provider is put on notice of the complaint, and has had the opportunity to conduct an internal review of the issue at an early stage. ● The anonymity of a complainant (A4.14) is not obviously a reasonable and proportionate reason for not engaging with a service provider (provided that the anonymity of the complainant can be preserved through use of redactions). It is important for services to have visibility over Ofcom's decision as to whether to commence an investigation, and to be provided with a copy of the initial assessment and consideration of the priority framework, as well as Ofcom's reasons for concluding that it should commence an investigation. This is fundamentally in the interests of fairness and transparency as part of Ofcom's public law duties, and in order to allow service providers the ability to seek legal recourse where appropriate. We would therefore ask that Ofcom amend A4.14 to provide that it will only be in exceptional circumstances that Ofcom may decide not to engage with a service provider during the initial assessment, such as an urgent or immediate risk of harm. **Confidentiality** We also note that at A4.12, Annex 11, the guidance envisages that Ofcom may request information from a service provider without using statutory information gathering powers. It would be helpful if the guidance could make clear that, even where information is provided voluntarily by a service provider (and not in response to a statutory requirement), the information will be treated as confidential by Ofcom and not shared with third parties. Furthermore, where the guidance generally refers to service providers being given advance notice of the publication of information relating to the enforcement proceedings (e.g. A4.24), we would ask that the guidance allows sufficient time for the service provider to make representations on confidentiality (e.g. 5 working days). **Confirmation Decisions and Financial penalties** The guidance outlines Ofcom's powers in relation to financial

Question (Volume 6)	Your response
	penalties for service providers in respect of contraventions. There is a sequencing point that could be clarified in the guidance around confirmation decisions and the imposition of financial penalties. We expect that financial penalties (and daily penalties) would not be payable until the deadline for appeal of a confirmation decision has expired and/or the final outcome of any appeal, however, it would be helpful if this could be clarified in the guidance. We note at A6.49, Annex 11, the guidance sets out Ofcom's enforcement powers relating to a service provider's breach of the duties relating to children's access assessments. However, where service providers self-certify that children are likely to access the service and accept that the child safety duties in the Act apply, it would be efficient if services are not required to undertake the formalities of conducting a children's access assessment. This section of the enforcement guidance should therefore clarify that it only applies where a service has failed to conduct a children's access assessment and not otherwise self-certified that it is possible for children to access the service and the "child user condition" is met. **Liability of related companies** In reference to the maximum penalty for group entities that have been found to be jointly and severally liable, A7.24 states that the relevant qualifying worldwide revenue consists of (i) the service provider and (ii) every other company that is in the same company group as the service provider. We assume that "same company group" refers to "group undertaking" (see Schedule 13, para 5(3)) as defined in section 1161(5) of the Companies Act 2006, and would suggest this is clarified by way of footnote. **Settlement procedure** A8.3 states that settlement is not equivalent to the type of discussions which take place between parties to litigation or potential litigation on a 'without prejudice basis'. It is unclear what this is intended to mean, as generally regulatory settlement discussions are conducted on a "without prejudice" basis, so that, if a binding settlement agreement is not concluded, the parties

Question (Volume 6)	Your response
	will not be permitted to refer to or seek to rely on any admissions, concessions, offers or proposals made in the course of settlement discussions. Otherwise, it leaves service providers with little incentive to enter into such a process. We raise the same point in relation to A8.33 which states that "*any additional documentary evidence provided during the settlement process would be placed on the case file and could be taken into account by Ofcom for the purposes of its final decision even if the settlement process is unsuccessful.*" We disagree that this information (obtained through a without prejudice process) should be provided to the final decision maker on the outcome of the regulatory process.

Question (Annex 13)	Your response
Question 54 (A13.1): Do you agree that our proposals as set out in Chapter 16 (reporting and complaints), and Chapter 10 and Annex 6 (record keeping) are likely to have positive, or more positive impacts on opportunities to use Welsh and treating Welsh no less favourably than English?	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.
Question 55 (A13.2): 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	**Confidential: N** We do not have specific comments on this question but would welcome further discussion with Ofcom if it would be helpful.

Question (Annex 13)	Your response
opportunities to use Welsh and treating Welsh no less favourably than English.	

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