Your response

Questions for all stakeholders	Your response
Question 19: What examples are there of effective use and implementation of any of the measures listed in article 28(b)(3) the AVMSD 2018? **The measures are terms and conditions, flagging and reporting mechanisms, age verification systems, rating systems, parental control systems, easy-to-access complaints functions, and the provision of media literacy measures and tools. Please provide evidence and specific examples to support your answer.**	Confidential? – N Some examples of good practice against these measures are listed below. A significant aspect is the combination of technology and human intervention – technology alone is insufficient. The question of how effective these measures are (in protecting users from harmful content), is discussed at Question 20. • **Flagging/reporting/complaining**: https://reportharmfulcontent.com/: third party reporting function that is a trusted flagger with a successful takedown rate of 91%. • **Age verification**: YOTI software – used in NSPCC and IWF partnership project for young people to self-report intimate images. • **Media literacy measures & tools**: There are increasing examples of social media companies which overlay untrustworthy content with links to trusted sources of information. There is a good novel text message training course by First Draft. This organisation also provides useful guidance about how best to flag inappropriate content, without drawing attention to it. • **Other - bespoke responses:** For example, You Tube switched off the comment function on videos depicting children, due to co-ordinated commenting activities by people with an unhealthy interest.
Question 20: What examples are there of measures which have fallen short of expectations regarding users' protection and why? **Please provide evidence to support your answer wherever possible.**	Confidential? – N People in Scotland (as across the whole of the UK) experience harms online, including when using major VSPS which already have the relevant protective measures in place (such as reporting mechanisms). Small scale research in Scottish schools suggest that many young people have accounts with VSPs, including children who are underage. Young people enjoy watching videos and livestreams , following a large and diverse range of influencers. Some young people broadcast their own videos and livestreams (and this may have increased due to the popularity of TikTok during lockdown). When asked about harms arising from all their online activities (VSP and other platforms), some young people in Scotland (primary and secondary) report skipping meals or missing sleep in

order to be online, and feeling pressure to do something online or spend money online. Some say they've believed something online that turned out to be false and some worry about things they have seen online, which can affect their sleep. A very small number have reported seeing upsetting content such as animal abuse and terrorist attack footage. In these extreme cases it is unclear if parental controls were in place, but a significant proportion of young people say there are no parental controls in use at home. And some young people reveal a sense of responsibility to 'toughen up' as if they should expect to see harmful content online. Engagement with professionals across Scotland reveals their concerns for the young people they work with, for example, young people who have been drawn into conspiracies such as flat earth (associated with video platform recommendation algorithm).

There are myriad examples of online harms affecting children and adults resulting from ineffective processes such as:

- Inadequate response when breaches of terms of use are reported to the platform (see Report Harmful Content service at Q 19, which operates only after a user has reported to the platform and received a dissatisfactory response).
- Harmful content which is not filtered out in services designed for children, or not caught by parental control filters (such gross violence within parody cartoon videos).
- Harmful content which is re-uploaded after removal (e.g. recent suicide video)
- Inappropriate takedowns (benign content which has been flagged as harmful).
- VSP responses to harmful content which are incremental reactions to negative publicity e.g. firstly remove specific pieces of content, then remove advertising placements, then tweak recommendation algorithm to stop promoting the content to others.
- Terms of use which do not meet ethical standards e.g. suppression of content from minority groups.

Regulation which ensures that VSPs not only implement but also consistently adhere to these measures will therefore be a considerable step forward. However, will compliance achieve the desired outcome of *protecting* users from harm? Challenges to effective regulation include adopting measures which are preventive and outcome-focused (see Q 26); and defining what is meant by harmful content (see Q 22).

	Systemic changes would be required to protect users from harmful content. A profit-driven business model promotes harmful content online due to user profiling, personalised advertising, personalised content recommendations and screentime targets. Users are not in control of the content they are served. In a competitive industry, VSPs may be more likely to rush new goods and services to market and less likely to assess impact, consider ethics, co-produce with users, or work jointly with competitors to improve safety.
Question 21: What indicators of potential harm should Ofcom be aware of as part of its ongoing monitoring and compliance activities on VSP services? Please provide evidence to support your answer wherever possible.	Confidential? – N Social media transparency reports (e.g. You Tube) set out some potential indicators of harm. However, care must be taken that outputs are not used as a measure of success or failure. For example, one service may appear more dangerous than another because it reports a larger number of illegal videos that have been removed. However, this is because it is proactively searching for such content. The number of takedowns is a blunt indicator, as there may be myriad reasons for content removal (including that a VSP is getting rid of controversial content). Similarly, there are issues around counting the number of reports, or complaints, depending on how these functions are marketed to users and how easy they are to use. There is a risk that targets act as perverse incentives – VSPs must be encouraged to report honestly. Similarly, indicators of potential harm must not be so restrictive as to hamper innovative approaches to user safety. It must be recognised that some VSP users are children. VSPs will be required to implement appropriate age verification measures and it would be anticipated that the Age Appropriate Design Code provides suitable guidance for adaption. VSPs which propose that there are no child users of their services should expect challenge on this issue. Indicators of harm should focus on encouraging system design which is protective of users. This would include not only measuring actual experiences of harm, but also harm prevention e.g. • % of all removed videos which occurred at the point of upload or within x timeframe; • % of successful appeals; • whether users in general feel able to upload any content they choose to a VSP (without fear of harassment, for example); • whether users know how to report a problem to the VSP, or feel confident that their problem would be addressed, etc.

	• Analysis broken down by equalities characteristics such as gender, age etc. Ongoing monitoring would not settle on maintaining baseline compliance but instead focus on planned improvement. It may be appropriate for Ofcom (or other suitable organisation) to conduct user research to provide a baseline for the above.
Question 22: The AVMSD 2018 requires VSPs to take appropriate measures to protect minors from content which 'may impair their physical, mental or moral development'. Which types of content do you consider relevant under this? Which measures do you consider most appropriate to protect minors? **Please provide evidence to support your answer wherever possible, including any age-related considerations.**	Confidential? – N There are existing models of harm which could apply here, such as the BBFC classifications, PEGI ratings and the UK Government's Online Harms white paper. Due consideration must firstly be given to illegal harms e.g. child sexual abuse imagery. However, relevant legislation includes devolved matters. For example the following Scottish legislation is not mirrored across the UK: • *Content which impairs a minor's physical, mental or moral development:* The Sexual Offences (Scotland) Act makes it illegal for anyone aged 16+ (not 18+) to send pornography to a child. • *Content inciting violence or hatred:* The Hate Crime and Public Order (Scotland) Bill proposes creating a new offence of stirring up hatred that applies to all hate crime characteristics identified in the Bill (not just racial hatred) and provides additional protections to freedom of expression. • *Content constituting criminal offences:* The act of threatening to share an intimate image has been illegal in Scotland since 2016 and cyberflashing (sending an unsolicited intimate image) has been illegal in Scotland since 2010. How would a UK-wide legislative and regulatory framework adequately take into account these differences in approach? Secondly, attention should be paid to those harms which children or adults define as having the greatest impact to children, as per research such as Ofcom (e.g. 12-15s most worried about bullying, abusive behaviour or threats (51%), viruses (46%), hate speech (42%) and content promoting self-harm (40%)). A significant risk is that VSPs state that children are prohibited from using their platforms, and therefore they do not need to provide protections for under-age users.

	Online risk is not the same to all. Vulnerable children are more at risk online than other children. Can/should VSPs identify vulnerable users and proactively provide extra protections? A final suggestion is that VSPs proactively educate current and potential users so they are better able to protect themselves. Education could include: • Transparency about the Directive, users' rights and how to complain; • Media literacy education so that users can make informed choices about how they use the service, effectively manage their privacy and security settings, and understand why they are served certain content and how to control that process.
Question 23: What challenges might VSP providers face in the practical and proportionate adoption of measures that Ofcom should be aware of? **We would be particularly interested in your reasoning of the factors relevant to the assessment of practicality and proportionality.**	Confidential? - N The pandemic has exposed some business continuity risk, whereby there were insufficient resources to maintain safety measures e.g. unable to respond to appeals. Also, effective measures require a combination of technology and human intervention – this is costly. A proportionate approach would be the expectation that human intervention is greater for larger organisations. The Age Appropriate Design Code guidance provides a proportionate approach to compliance (for example the age-appropriate application), which could similarly be adopted here.
Question 24: How should VSPs balance their users' rights to freedom of expression, and what metrics should they use to monitor this? What role do you see for a regulator?	Confidential? – N The Human Rights Act (article 10) protects people's rights to hold their own opinions and express them freely without government interference. Online services potentially provide a platform to support freedom of expression, particularly in relation to anonymity. However, online services can invoke a 'chilling effect' whereby users refrain from sharing or commenting due to abuse they might expect. It should be noted that VSPs are private, not public services and they are not 'a public square' so do not have the same responsibilities as a public service provider. VSPs define appropriate behaviour within their terms of use, which users must accept if they wish to use the platform. The role of a regulator is therefore somewhat limited. Regarding metrics, there may be appropriate proxy measures for freedom of expression, such as whether minority group users say they have avoided sharing due to potential abuse on a certain platform. One aspect to consider is the idea of 'freedom of speech vs freedom of reach'. A VSP may allow certain content on its platform, but should not actively promote and spread this

	content (e.g. via recommendation algorithms) where this would be harmful to other users.
Question 25: How should VSPs provide for an out of court redress mechanism for the impartial settlement of disputes between users and VSP providers? (see paragraph 2.32 and article 28(b)(7) in annex 5). **Please provide evidence or analysis to support your answer wherever possible, including consideration on how this requirement could be met in an effective and proportionate way.**	Confidential? - N The volume of potential disputes is a significant challenge to resource. As previously mentioned, relevant legislation includes devolved matters. This is further complicated by terms of use. For example, a VSP could adopt the strongest form of all pieces of UK legislation into its terms of use. This cannot be challenged through the out-of-court redress mechanism as these are simply the terms of service, to which a user has consented. There may be a challenge to Ofcom and VSPs operating in the UK in the public's understanding of the scope of this legislation (i.e. which companies fall within Ofcom's regulatory scope). The route of redress for disputes with all companies operating in the UK should be made clear.
Question 26: How might Ofcom best support VSPs to continue to innovate to keep users safe?	Confidential? - N Many of the measures are reactive rather than preventive (e.g. removing content after it has been uploaded, viewed and shared). To effectively protect users from experiencing harmful content requires changing focus towards prevention. This could include the following: • For specific pieces of harmful content, finding innovative ways to prevent the initial harm. For example, detecting (and preventing) the content during initial upload, using industry alerts for a co-ordinated immediate response, or introducing a time delay into a livestream. Innovate safety measures in response to new threats (as users find ways to circumvent existing measures). • Design safer services such as safety settings switched on by default. Use approaches such as contextual safeguarding and behavioural nudge techniques to design online services and community spaces which are protective and inclusive by default. Identify vulnerable users to put in place extra protections. Design out processes that increase risk e.g. recommending harmful content because it is more profitable, or targeting harmful content to certain users who are profiled to have an interest in it. Educate users so that they can protect themselves. A VSP's ability to innovate in response to the online threat landscape is critical. However, there are technical, commercial

	and financial challenges to doing so. VSPs may profit more from harmful than benign content. Regulation can encourage safety innovation by: • Requiring VSPs (proportionally) to demonstrate preventive approaches to achieving the AVMSD measures; • Using indicators of harm which are flexible enough to enable VSPs to pilot innovative approaches e.g. to test a theory of change; and • Supporting VSPs to collaborate and share best practice in terms of prevention.
Question 27: How can Ofcom best support businesses to comply with the new requirements?	Confidential? - N Ensure clear separation of support and regulatory services, so VSPs feel confident to ask for support without fear of penalty. Communicate the new requirements to the public, so that users understand their rights.
Question 28: Do you have any views on the set of principles set out in paragraph 2.49 (protection and assurance, freedom of expression, adaptability over time, transparency, robust enforcement, independence and proportionality), and balancing the tensions that may sometimes occur between them?	Confidential? – N Some of these principles have been addressed in earlier questions in this document. Regarding protection and assurance, key challenges lie in ensuring a clear definition of online harms and giving due consideration to different legislation across the UK. Also, as discussed earlier, protection necessarily involves a balance of preventive and reactive measures. Regulation must focus on the outcome as well as process, to ensure measures are effective. Although the number of companies involved is small, it is likely that whatever regulatory processes are put in place will set the standard to which future online harms regulation adheres. The risk is that 'successful' regulation is interpreted as companies achieving the minimum (such as responding to complaints timeously). This could lose sight of the systemic changes needed to deliver the outcome of protecting users.