Your response

Questions for industry	Your response
Question 1: Are you providing a UK-established service that is likely to meet the AVMSD definition of a VSP? **Please provide details of the service where relevant. The establishment criteria under the AVMSD are set out in annex 5.**	Confidential? – Y / N
Question 2: Is your service able to identify users based in specific countries and do you provide customised User Interfaces (UI), User Experience (UX) functionality or interaction based on perceived age and location of users?	Confidential? – Y / N
Question 3: How does your service develop and enforce policies for what is and is not acceptable on your service? (including through Ts&Cs, community standards, and acceptable use policies) **In particular, please provide information explaining:** • **what these policies are and whether they cover the categories of harm listed in the AVMSD (protection of minors, incitement to hatred and violence, and content constituting a criminal offence – specifically Child Sexual Exploitation and Abuse, terrorist material, racism and xenophobia);** • **how your service assesses the risk of harm to its users;** • **how users of the service are made aware of Ts&Cs and acceptable use policies; and** • **how you test user awareness and engagement with Ts&Cs.**	Confidential? – Y / N
Question 4: How are your Ts&Cs (or community standards/ acceptable use policies) implemented?	Confidential? – Y / N

In particular, please provide information explaining: • **what systems are in place to identify harmful content or content that may breach your standards and whether these operate on a proactive (e.g. active monitoring of content) or reactive (e.g. in response to reports or flags) basis;** • **the role of human and automated processes and content moderation systems; and** • **how you assess the effectiveness and impact of these mechanisms/ processes.**	
Question 5: Does your service have advertising rules? **In particular, please provide information about any advertising rules your platform has, whether they cover the areas in the AVMS Directive, and how these are enforced. See Annex 5 for a copy of the AVMSD provisions.**	Confidential? – Y / N
Question 6: How far is advertising that appears on your service under your direct control, i.e. marketed, sold or arranged by the platform? **Please provide details of how advertising is marketed, sold and arranged to illustrate your answer.**	Confidential? – Y / N
Question 7: What mechanisms do you have in place to establish whether videos uploaded by users contain advertising, and how are these mechanisms designed, enforced, and assessed for effectiveness?	Confidential? – Y / N
Question 8: Does your service have any reporting or flagging mechanisms in place (human or automated)? **In particular, please provide information explaining:**	Confidential? – Y / N

• **what the mechanisms entail and how they are designed;** • **how users are made aware of reporting and flagging mechanisms;** • **how you test user awareness and engagement with these mechanisms;** • **how these mechanisms lead to further action, and what are the set of actions taken based on the reported harm;** • **how services check that any action taken is proportionate and takes into account Article 10 of the European Convention of Human Rights (freedom of expression);** • **how users (and content creators) are informed as to whether any action has been taken as a result of material they or others have reported or flagged;** • **whether there is any mechanism for users (including uploaders) to dispute the outcome of any decision regarding content that has been reported or flagged; and** • **any relevant statistics in relation to internal or external KPIs or targets for response.**	
Question 9: Does your service allow users to rate different types of content on your platform? **Please provide details of any rating system and what happens as a result of viewer ratings.**	Confidential? – Y / N
Question 10: Does your service use any age assurance or age verification tools or related technologies to verify the age of users? **In particular, please provide information explaining:** • **how your age assurance policies have been developed and what age group(s) they are intended to protect;** • **how these are implemented and enforced;** • **how these are assessed for effectiveness or impact; and** • **if the service is tailored to meet age-appropriate needs (for example, by**	Confidential? – Y / N

restricting specific content to specific users), how this works.	
Question 11: Does your service have any parental control mechanisms in place? **In particular, please provide information explaining:** • **how these tools have been developed;** • **what restrictions they allow;** • **how widely they are used; and** • **how users of the service, and parents/ guardians if not users themselves, are made aware of and encouraged to use the parental control mechanisms that are available.**	Confidential? – Y / N
Question 12: Does your service have a complaints mechanism in place? Please describe this, including how users of your service can access it and what types of complaint they can make. **In particular, please provide information explaining:** • **any time limits for dealing with complaints;** • **how complainants are informed about the outcomes of complaints;** • **any appeals processes, how they work, and whether they are independent from the complaints processes; and** • **the proportion of complaints which get disputed or appealed.**	Confidential? – Y / N
Question 13: What media literacy tools and measures are available on your service? **In particular, please provide any relevant information about:** • **how you raise awareness of media literacy tools and measures on your service;** • **how you assess the effectiveness of any media literacy tools and measures provided on your service; and** • **how media literacy considerations, such as your users' ability to understand and respond to the**	Confidential? – Y / N

content available to them feature in your thinking about how you design and deliver your services, for example in the user interfaces, flagging content and use of nudges.	
Question 14: Do you publish transparency reports with information about user safety metrics? **Please provide any specific evidence and examples of reports, information around the categorisation and measurements used for internal and external reporting purposes, and whether you have measures in place to report at country/ regional level and track performance over time.**	Confidential? – Y / N
Question 15: What processes and procedures do you have in place to measure the impact and effectiveness of safety tools or protection measures? **If not already captured elsewhere in your response, please provide information relevant to all of the measures listed above explaining:** • **how you test and review user awareness and engagement with each measure (including any analysis or research that you would be willing to share with Ofcom);** • **how often policies and protection measures are reviewed, and what triggers a review; and** • **how you test the impact of policies on users and the business more generally, such as how you balance the costs and benefits of new tools.**	Confidential? – Y / N
Question 16: How do you assess and mitigate the risk of inadvertent removal of legal or non-harmful content? **In particular, please provide any information on:** • **how freedom of expression is taken into account during this assessment;** • **how appeals are handled and what proportion are successful; and**	Confidential? – Y / N

• **audits of automated removal systems and, if you have them, any metrics that relate to their effectiveness.**	
Question 17: Have you previously implemented any measures which have fallen short of expectations and what was your response to this? **Please provide evidence to support your answer wherever possible.**	Confidential? – Y / N
Question 18: How does your service develop expertise and train staff around different types of harm? (e.g. do you have any partnerships in place?)	Confidential? – Y / N

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.**	The South West Grid for Learning's (SWGfl) 3 helplines are examples of easy to access support services with mediatory capabilities. They embrace collaborative working with industry to seek resolution for their respective clients and ultimately to remove harmful/ illegal content online. SWGfL have been providing helpline services since 2011, when the UK Safer Internet Centre launched its Professionals Online Safety Helpline (POSH) supporting those working with children across the UK. In 2015, this was joined with the launch of the Revenge Porn Helpline (RP helpline) to coincide with new legislation. The RP Helpline supports victims who are facing intimate image abuse (IIA) - having their intimate images shared without consent. SWGfL launched Report Harmful Content (RHC) in December 2019 to support victims facing legal but harmful online content. RHC was developed over a 5-year period and following a pilot phase beginning in December 2018 launched officially in 2019.

	RHC provides information, support and mediation to all UK users (over 13 years old) with regards to legal but harmful online content. At a basic level, the service provides definitions of what legal but harmful online content is; support for users facing these issues, and direction of how to report these issues to social media and online service providers. If a user has reported their issue and has received a null or an unsatisfactory (from their perspective) response from the social media or online provider, RHC will assess the report. In assessing the report, RHC has the opportunity to understand the context in order to determine if the response was unfair. If the response was fair, the user will be provided with advice and an explanation. If the conclusion is that the response was unfair, RHC might either provide further direction or accept the report and represent the claimant with the online provider. SWGfL has developed in-depth understanding of terms and conditions and community standards to enable it to adequately represent claimants. In the 9 months since officially launching RHC, of the reports accepted, 91% have resulted in the harmful content being removed. We would recommend and best practice shows that the use of independent organisations and reporting processes for complaints/ appeals functions when internal reporting routes through VSP's have been exhausted, such as the role NetSafe have in NZ. For example, in the immediate aftermath of the Christchurch attacks, whilst New Zealand Law Enforcement and Government departments worked to identify content to be removed, the operational relationships established by NetSafe (as an independent NGO with a regulatory role) provided capacity and greatly expedited the efficient removal of identified harmful content. This is an important aspect and SWGfL recommend that services, such as RHC can adopt this role working alongside the regulator. Examples of good practice across industry include but are not limited to: • Processes that prevent content that is considered in scope from appearing on platforms in the first place, e.g. the

	Facebook pilot project in conjunction with the RP helpline which hashes intimate imagery that has been threatened to be shared non-consensually to prevent it from being uploaded to Facebook/ Instagram. • AI being used as a tool for good and to enhance safety by design principles. For example, the gamification of safety features to appeal to younger users such TikTok's safety account. • Systems and processes promoting the wellbeing of users for example responding to reports made about self-harm/ suicide/ eating disorder content signposting to sources of support and Snap's Here for You Initiative and Roblox's Digital Civility & Safety Mission. • Anti-bullying controls that encourage user management of incidences. For example, managing/ turning off comments on Instagram, Mute functionality on Twitter.
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? - Y
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.**	It is advantageous that Ofcom have a comprehensive understanding of the complexities surrounding online harms and we would recommend that they continue to consult and work alongside industry, NGO's, and frontline support services such as SWGfL's helplines to ensure awareness of online behaviours indicative of harm remains up to date. Academic research and annual reports in the sphere of online safety provide up to date information about the indicators of harm being experienced and reported and we would recommend drawing on this knowledge to help inform regulatory powers. RHC has released its

	first annual report highlighting the trends of harm being experienced by certain users: https://swgfl.org.uk/research/report-harmful-content-annual-report-2020/. Similarly resources such as the Online Resillience Tool: https://www.headstartkernow.org.uk/digital-resilience/ (created by Headstart Kernow in consultation with SWGfL's POSH) could be really useful for VSP's ion scope to help them detect behaviour that may indicate harm on their platforms.
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.**	Graphic violent content, pornographic content, content glorifying/ promoting mental health disorders such as self-harm, suicide content and eating disorders, illegal content (e.g. CSAM, Intimate Image Abuse, terrorist content), content promoting the sale/ purchasing of illegal goods, videos that contain solely text based narratives containing illegal themes (e.g. CSA/CSE narratives created for the purpose of sexual gratification). Where imagery is apparent this should include CGI imagery as well as deep fakes/ photo-shopped imagery and any other edited imagery so long as it is seen to fall into the categories above. Age Verification has the potential to be a game changer in this space and if it can be made to work effectively, it would make sense to deploy this across all VSP's in scope to allow them to prevent underage users from accessing their platforms and to age gate legal but harmful content. However as a lot of the content detailed above is already classed as illegal, we would recommend firmer measures to prevent the content from being readily available to anyone in the UK regardless of age. SWGfL and the UK Safer Internet Centre partners are well placed to provide practical guidance to organisations about safety by design with regard to young people. The difficulty here is clearly that, whilst services in scope of this directive will already be taking preventative actions in line with their own community standards or by adhering to this directive in the future (e.g. by being active members of the IWF etc), it is independent sites, forums, discussion boards and adult content sites where the biggest problem lies.

	For the most part, these are hosted outside the UK and not in scope. SWGfL have a unique insight into the places where this content is strife and would recommend that Ofcom work with SWGfL to ensure that VSP's in scope can prevent harmful and in the respect of IIA, illegal content from appearing on their platforms, expanding on the services already provided by the CITRU and IWF. We would recommend making it a requirement for all VSP's in scope to be members of such initiatives. Assuming the continuation of this type of function and given that it is in support of VSP's and Ofcom, sufficient funding will be required to sustainably finance this and similar operations.
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.**	Economies of scale have the potential to cause an issue here: 1) For smaller VSP's in scope who may simply not have the taskforce or technology available to put in place measures decided on. 2) When considering how VSP's are adhering to the directive e.g. for removal of content, transparency data will show large numbers of content removed for a well-established VSP with a huge user base, compared to tiny figures for a startup/ smaller scale VSP but this is relative. To this end, we would recommend that any VSP's in scope share best practice and where possible make technology enabled solutions open source so that all can benefit. There may also be benefit in paring smaller VSP's and start-ups with larger scale VSP's so they can benefit from each other's expertise and share/ model safety by design principles in respect of adhering to the directive. Penalties imposed for not adhering to measures will need to be proportional to the scale/ size of the VSP.

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?	SWGfL recommends that where the balance of freedom of expression and user's rights is bought into question, incidences be reviewed on a case by case basis as the one size fits all approach does not work here. When making such decisions we recommend VSP's in scope work to the principle that whilst public discourse is important, a user's safety and wellbeing should be deemed more so (i.e. where credible threats are made towards identifiable targets (people, groups etc) or a campaign of harassment/ pile-on abuse ensues against said target(s), user rights should prevail).
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.**	Most of the VSP's in scope will already have a user reporting/ flagging system in place internal appeals process when a user report has not been actioned. We would recommend that all routes to redress are easy to navigate and discover and that these are not just limited to users of the service (i.e. users should not have to sign in to be able to make reports). SWGfL would consider all its helplines to be unique in the UK and paramount to the user redress process, but of particular relevance, here are RHC; providing users and victims of legal but harmful online content with independent support and redress and RP Helpline; providing victims of IIA with support and redress. SWGfL has developed a good understanding of industry policies and the law relating to online criminal behaviour and will escalate content for removal only when we know it breaches community standards and/ or the law. We would recommend appointing designated bodies as super complainants as outlined in the online harms white paper to enable super complaints to be bought to the regulator when all of the above routes to redress have failed. Given the remit of the helplines SWGfL operates as outlined above, we recommend Ofcom works with SWGfL to help fulfil this obligation. RHC is a primary example of a designated body that could bring 'super complaints' to a

	regulator. Victims would benefit from having strengthened their complaint and bring some further redress for what is likely to be a distressing situation. In terms of language and for clarity, SWGfL considers that users should 'report' issues to the VSP in scope. If they are unsatisfied with the response they receive they submit this to RHC and if RHC agrees (against criteria), a 'complaint' is then submitted to the social media or online provider. If the outcome of this is unsatisfactory, RHC submit a 'super complaint' to the regulator. Assuming the continuation of this type of complaint and super complaint function and given that it is in support of VSP's and Ofcom, sufficient funding will be required to sustainably finance this and similar operations.
Question 26: How might Ofcom best support VSPs to continue to innovate to keep users safe?	There are a number of areas where VSP's need practical guidance in order to help innovation: Age verification as detailed in Q22. • Violence against Women and Girls (VAWG) - It is apparent from our work across all SWGfL helplines that VAWG accounts for a disproportionately large amount of the cases we deal with and we propose that guidance is issued in this area specifically. SWGfL helplines would be keen to collaborate with other stakeholders to produce this. • SWGfL would recommend support for VSP's developing products that accept or host digital images and video to be required in order to prevent illegal imagery from being uploaded by using hashing technologies. For example, SWGfL would like to see all VSP's in scope utilise the IWF hash list to prevent known illegal CSAM images from being uploaded. • The area of extremism and radicalisation is complex and more support from Government, academics and experts need to be given to providers, particularly with regards to definitions, search terms and illegal imagery identification.

	• SWGfL have a unique insight into the places where the content outlined in Q22 is strife and would recommend that Ofcom work with SWGfL to ensure that VSP's in scope have the ability prevent harmful and, in the respect of IIA, illegal content from appearing on their platforms, expanding on the services already provided by the CITRU and IWF.
Question 27: How can Ofcom best support businesses to comply with the new requirements?	SWGfL would suggest care in constructing reporting indicators, especially numerical indicators. Numerical indicators, alone, may be open to misinterpretation and would require context. For example, a simple rise in reports may both highlight an inherent issue and it may also represent a rise in confidence of users to be making reports. Care is also needed to avoid influencing providers to make amends to their reporting policies or processes that aim to improve ratings rather than serve the interests of users. Again, as an example, a measurement of report closure rates - providers may choose to focus on closing reports more rapidly to increase their performance, however this may be to the detriment of their users and user experience. SWGfL would like to also take this opportunity to highlight that VSP's should be required to report on the deployment of the IWF URL/Hash list. SWGfL would encourage all providers to be members of IWF and to deploy their services across their infrastructure. This would highlight those providers who are either not members and/or not protecting their users from illegal online child sexual abuse materials. To support users, particularly schools, in understanding if their filtering or ISP provider protects their connection from websites identified by IWF as containing illegal child sexual abuse material, SWGfL has developed an online utility - http://testfiltering.com/ The utility tests for the deployment of both the IWF and CTIRU URL lists and presents the user with a pass/fail result. The utility is connected with the definitions of 'appropriate filtering and monitoring' published by SWGfL to empower users.

	In our experience of working with industry, trust is the key to mutually beneficial relationships and effective transparency. It's vitally important that users of the VSP's in scope trust the providers and, in order to gain this trust, transparency with users is paramount. Any transparency reports created should be viewable by all and not just available to the government/ regulatory body. This will help to build trust in the wake of revelations about user data leaks and non-compliance with GDPR and the DPA. Involving industry meaningfully will help accountability, as this will allow time to establish commitments from as to what they can be held accountable for. Ultimately, being seen to be accountable increases trust.
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?	It is great to see the acknowledgment of Ofcom to interact with the ICO's Age Appropriate Design Code and the CMA as well as a commitment to continue to work with regulatory counterparts in other countries. SWGfL works in collaboration with a number of international partners. For example, SWGfL supported the Australian eSafety Commissioner's office in establishing their helpline to support victims of IIA in Australia RHC has benefitted from support from colleagues in both Australia, but primarily the NGO NetSafe in New Zealand. The Harmful Digital Communications Act was introduced in 2016 in New Zealand and introduced civil and legal definitions of harmful digital content, together with a regulatory role for NetSafe in assessing cases of Harmful Digital Content. RHC greatly benefited from the understanding NetSafe's experience of establishing this service. We would recommend continuing to consult with international counterparts and SWGfL in order to understand and draw upon their experiences of balancing tensions that occur between principles such as those outlined in the call for evidence.