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.**	ARTICLE 19 has not specifically monitored the implementation of the measures listed in article 28 (b) (3) of the AVMSD 2018. We note, however, that major platforms such as YouTube have long had terms and conditions and used measures such as flagging, content removal and reporting mechanisms. Whether or not these measures are considered 'effective' very much depends on the definition of effectiveness and how they are assessed. By and large, effectiveness has been assessed by reference to the volume of content taken down. In our view, this is mistake. Whether or not companies report a significant volume of takedowns is also a function of how they write their terms of service: if the definition of 'harmful' content is expanded, it is more likely that the volume of removed content will go up. It is also more likely that legitimate content will be removed. It could also lead to the removal of content of particular groups leading to claims of bias (see e.g: Amleh, Are YouTube's policies biased against Palestinians, April 2020: https://7amleh.org/storage/Youtube_0420_En glish%20(4).pdf). In those circumstances, it is unclear whether the measures listed would be considered 'effective'. We also draw attention to ARTICLE 19's Missing Voices Campaign, which presents stories of individuals whose content has been removed without proper due process. One of the key criticisms of companies' appeals mechanisms is the lack of notification or reasons being given for the removal. https://www.article19.org/campaigns/missingv oices/. Given the volume of content flagged by filters, it is perhaps unsurprising but it is deeply unsatisfactory. Whilst there is currently insufficient conclusive data about appeals mechanisms, anecdotal data suggests that they are not 'effective' and in any event lack due process safeguards.
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.**	ARTICLE 19 believes that the following measures have fallen short: - Transparency reporting - Companies' complaint mechanisms - Flagging system - Use of filters leading to over-removal of content

Transparency reporting and complaint mechanisms:

There is still very limited information available about the scale of wrongful removals of content, particularly as it relates to particular regions or countries. The only information available stems from broad figures about successful appeals against takedowns. For instance, in its Jan-March 2020 Transparency report, Facebook reported the following:

- Content reinstated on appeal has decreased for nudity -related content: whereas nearly 4 million pieces of content had been appealed in April-June 2019, it was down to about 2.3 million in January-March 2020. The number of pieces of content restored has also decreased from just over one million in April-June 2019 to just over 600,000 in January-March 2020.

- The number of appeals on 'hate speech' grounds remains stable, hovering around 1.3 million between January-March 2019 and the same period in 2020. Since a peak of around 170,000 pieces of content restored in July-September 2019, the number of reinstated pieces of content has steadily decreased to just over 60,000.

- Appeals against decisions on organised hate have increased but not led to a significant amount of reinstated content. Appeals against actioned content on grounds of organised hate have increased since October-December 2019 reaching just over 230,000 in January-March 2020. The amount of restored content is low, at about 50,000 pieces of content. Some content is restored automatically, primarily terrorist-related content with nearly 300,000 pieces of content restored without appeal.

As noted above, our Missing Voices Campaign has collected stories of wrongful removals of content on the basis of companies' community standards. There are numerous examples of journalists, artists, human rights defenders and marginalised groups experiencing the wrongful removal of content. This undoubtedly has an impact on their fundamental rights and ability

to do their job effectively, including by holding governments and others to account. More information about the Missing Voices Campaign is available from here: https://www.article19.org/campaigns/missingvoices/

Flagging mechanisms
It remains highly unclear who are trusted flaggers for given types of content, what criteria are being used to select them, whether they include government authorities and if so which ones, and what the implications are of obtaining trusted flagger status. At the very least, more information should be provided about these programmes. In particular, specific data should be provided about the number of pieces of content removed on the basis of trusted flaggers' reporting, whether any such content is appealed and the extent to which such appeals are successful and content reinstated.

Use of filters
ARTICLE 19 opposes the use of mandatory filters. We note, however, that filters are used in practice to flag content. Reliance on filters has also accelerated throughout the pandemic leading to over removal of content. (See e.eg. Techcrunch, *YouTube warns of increased video removals during Covid-19 crisis,* 16 March 2020*: https://techcrunch.com/2020/03/16/youtube-warns-of-increased-video-removals-during-covid-19-crisis/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAFD6SP9E0iVywRNFfi8Eurph93iIDfWTSjoWWBiq30_YqubfgpK1shzktWPDJfLbtXPBDf0okzhCdJNQywv0oJnjlzgKPeGAGHX2xvcLjQk3FtbTiQRwZ6vY990hWmoFlvwBrCr9j0pOVDNvKMEpqbUg2a-TJdTpJCvO3EtqvuNX*

The latest YouTube enforcement of community standards report (April-June 2020) suggests that the use of filters has led to almost double the amount of content being removed, from about 6 million videos being removed in Jan-March 2020 to over 11 million in April-June 2020. The BBC recently reported that about 50% of videos are reinstated on appeal (see here:

	https://www.bbc.co.uk/news/technology-53918584), although this information is not readily apparent in YouTube's enforcement report. In any event, it suggests that the accuracy of filters continues to be wanting. Yet, very little information is available about the use of filters, how they are designed or their error rate (false positives and false negatives).
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.**	ARTICLE 19 believes that it is impossible to answer this question without first defining what 'harm' means. It is unclear what content may be considered 'harmful', or by reference to whom. For instance, VSPs could potentially create 'harm' to users' by violating data protection rules while collecting or processing their data or by unduly removing the content they want to share on the platform. They could also create harm to business users if platforms deal with them in an unfair manner. We also warn against the use of terms such as 'potential' as opposed to 'actual' harm. It only highlights that the regulator has in its sight undefined 'harms' that are also highly speculative. There is, for example, little agreement on what constitutes 'violent content,' still less that it causes 'harm'. The same is true of 'disinformation', for instance. Like the government in its White Paper, Ofcom seems to elide content which might be seen as undesirable, on the one hand, with the idea that such content is 'harmful.' It should be obvious to any reasonable person that it is not necessary for something to be 'harmful' in order to be undesirable, and equally obvious that the law permits people to do many things that might be viewed by others as undesirable.

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.**	ARTICLE 19 notes that the AVMSD 2018 expressly mandates online video-sharing platforms services to put in place 'appropriate' measures, such as age verification and parental control systems, to protect children from material that may impact their physical, mental or moral development. ARTICLE 19 understands the concerns of child protection organisations around the availability of 'harmful' material online and their potentially negative impact on the development of children. Nonetheless, we are worried that proposals for greater regulation in this area could entrust a regulator such as Ofcom with powers to decide what amounts to 'harmful' content online in the absence of primary legislation to that effect. For instance, it is highly unclear what 'self-harm' means or what form it might take, e.g. whether it includes websites about anorexia, alcoholism, drug taking or dangerous sex. Moreover, while the idea of removing 'self-harming' websites may sound attractive, in practice, educational websites about this issue may end up being caught in filters that are currently unable to capture such nuance (see, for example a joint submission of ARTICLE 19 and Prostasia foundation, https://prostasia.org/wp-content/uploads/2019/12/Prostasia-case-and-Paypal-submission-December-2019.pdf). Equally, youth who visit self-harm websites may be joining online groups to share experiences. Seek support and connect with others. Although some of these conversations may be unhealthy, others may not be. Shutting down such websites could therefore have a detrimental impact on such youth looking out for a sense of community and belonging. ARTICLE 19 is also concerned about mandatory technical solutions, such as proactive filtering or age-verification systems that could have a disproportionate impact on the rights to privacy and freedom of expression of both children and

adult Internet users. We are equally concerned about any proposals that could both undermine encryption and/or online anonymity. In our view, these are cornerstones of the protection of freedom of expression, privacy and information security online and should be strongly protected.

Instead, we believe that social media companies should continue to adopt measures such as content rating and parental control systems on a voluntary basis. They should also be more transparent about their content moderation practices and provide complaints mechanisms for wrongful removal of content or for when they refuse to take content down. In addition, they should contribute to media literacy efforts for both parents and children. (For details of ARTICLE 19's views on preventing 'online harms' for children, see Malcolm and Guillemin, *Internet companies alone can't prevent online harms,* April 2020: https://www.article19.org/resources/internet-companies-alone-cant-prevent-online-harms/)

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.

VSP providers face a number of challenges in adopting practical yet proportionate measures to balance users' right to freedom of expression with other interests:

1. Scale. However, we note that this cannot be solved simply by increasing the use of AI, which has obvious limits. Indeed, AI itself is part of the challenge since it is inherently incapable of making nuanced assessments about context in a range of situations, but particularly as it regards 'hate speech' or 'violent extremism'. In practice, this means that companies remove content first by default and that any wrongful removal is fixed after the fact through appeals mechanisms. However, this turns the principles of protecting freedom of expression and due process on their head. Content should only be removed *after* an assessment has made of its legality by a court or other independent

adjudicatory body. Similarly, removal on the basis of community standards should take place *after* the uploader has been given an opportunity to respond to any complaint about his or her content.

At the very least, the use of technical tools should be far more transparent, recognising and accounting for AI's inherent inability to understand context and rhetoric truths: they can only solve part of the problem. For this reason, it is vital to have a 'human in the loop' in order to make more nuanced judgments about the compatibility of content with community standards.

2. VSPs' incentives to provide better tools for content moderation could be eliminated by the threat of losing immunity from liability for third-party content. A balance should be found between the need to guarantee that platforms act responsibly and the need to incentivise innovative solutions for content moderation challenges. One way to do so is to focus on VSPs complying with transparency and due process requirements rather than focusing on content itself and removal metrics.

3. The risk of running into liability, coupled with the massive use of AI tools for content moderation, creates incentives and conditions for over-removal by VSPs.

In assessing proportionality, we would invite Ofcom to consider the case-law of the European Court of Human Rights on website blocking orders. In Kharitonov v Russia, the Court noted that the law did not require the government agency (*Roskomnadzor*) to check whether the IP address used by the website targeted by a website blocking order was used by more than one website or to establish the need for blocking by IP address. It found that that manner of proceeding could, and did in the *Kharitonov v. Russia*, have the practical effect of extending the scope of the blocking order far beyond the illegal content which had been originally targeted. Both the original determination and Roskomnadzor's implementing orders had been made without any advance notification to the parties whose rights and interests were likely to be affected. The blocking measures had not been sanctioned by a court or other independent

adjudicatory body providing a forum in which the interested parties could have been heard. Nor did the Russian law call for any impact assessment of the blocking measure prior to its implementation. The Court concluded that there had been a breach of Article 10 ECHR. For more information about these cases, please see ARTICLE 19's press release, available here: https://www.article19.org/resources/russia-european-court-judgment-is-victory-for-freedom-of-expression/)

In ARTICLE 19's view, the Court's case-law points strongly in the direction of requiring human rights impact assessments in relation to content moderation systems, including filters and whether they lead to over-removal of content.

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?	ARTICLE 19 notes that once again, it is difficult to understand the question. To begin with, it is unclear against what VSPs should balance their users' rights to freedom of expression. ARTICLE 19 notes that a similar approach was envisioned in an earlier version of the French Avia Bill, which required the regulator to assess the extent to which freedom of expression was respected by platforms (i.e. whether content was 'excessively removed' by platforms):https://www.article19.org/resources/france-analysis-of-draft-hate-speech-bill/ In our view, this is difficult to do because community standards fall below international standards on freedom of expression. The main indicator of wrongful removal is the number of successful appeals. However, it is unclear if these are available against decisions made on the basis of filters and how users can argue

against the decision that was made without being given any reasons for it. Moreover, the incentives for users to use appeals mechanisms are inexistent. It creates friction and in practice, it seems highly unlikely that users will seek to challenge, e.g. the wrongful removal of 'terrorist' content for instance. Therefore, it is almost impossible to know how much legitimate content is removed.

Some over-removal issues may be addressed through human rights impact assessments of filters and the extent to which filters are biased and may remove content from particular groups. Greater transparency would at least help establish the scale of the problem when it comes to unduly zealous enforcement of community standards.

Finally, we note that VSPs should respect the UN Guiding Principles on Business and Human Rights. These provide a series of steps that businesses can take to mitigate the human rights impact of their business activities. We further note that the Office of the High Commissioner for Human Rights recently released a report on enhancing the effectiveness of non-State-based grievance mechanisms in cases of business-related human rights abuses. The report contains useful recommendations about potential benchmarks for assessing VSP's internal complaints' mechanisms:
https://www.ohchr.org/EN/Issues/Business/Pages/ARP_III.aspx

Question 25: How should VSPs provide for an out of court redress mechanism for the impartial settlement of disputes between	ARTICLE 19 suggests that VSPs and stakeholders should engage into the creation of Social Media Councils (SMCs) at the national level. The UN Special Rapporteur on the Promotion and

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.**	Protection of the Right to Freedom of Opinion and Expression endorsed this suggestion when he recommended in his April 2018 report that "all segments of the ICT sector that moderate content or act as gatekeepers should make the development of industry-wide accountability mechanisms (such as a social media council) a top priority" (https://undocs.org/pdf?symbol=en/A/HRC/38/35) In our view, the SMC would provide a forum where 'appropriate measures' under Art. 28b can be discussed, fine-tuned, assessed or reviewed by representatives of VSPs and all stakeholders. When looking at the best approach to the societal challenges related to content moderation, Internet companies cannot be expected or even encouraged to take the place of sex educators, therapists, social workers, researchers, media literacy experts, journalists and other voices in society. As it enables a broad participation from business and civil society, the SMC serves to elaborate a common understanding not only of which type of content should be moderated but also of the appropriate and realistic technical approaches to moderation. The SMC would also serve as an appeals mechanism: users will have access to an independent, external body that can make decisions on disputes related to content moderation. ARTICLE 19 has developed the idea of Social Media Council as a model for a multi-stakeholder accountability mechanism that provides a transparent, independent and accountable forum to address content moderation issues on social media platforms on the basis of international standards on human rights. We have gathered comments and suggestions on the SMC through numerous discussions with international experts and CSOs and an online public consultation.[1] The concept of SMC has integrated the current

---

[1] The consultation background document is available at https://www.article19.org/social-media-councils/. In February 2019, together with the Special Rapporteur on freedom of expression and the Global Digital Policy incubator at Stanford University, ARTICLE 19 brought together academics, civil society organisations and dominant social media platforms (Twitter, Facebook, YouTube) for a 2-day discussion of the concept: the report from the conference is available at https://cddrl.fsi.stanford.edu/global-digital-policy-incubator//content/social-media-councils-concept-reality-conference-report. In 2019, the SMC was further discussed during events such as UNESCO's World Press Freedom Day, RightsCon Tunisia, a workshop of the Organisation Internationale de la Francophonie or the Council of Europe's World Forum for Democracy.

international academic and policy debates on the future of the regulation of social media platforms.[2]
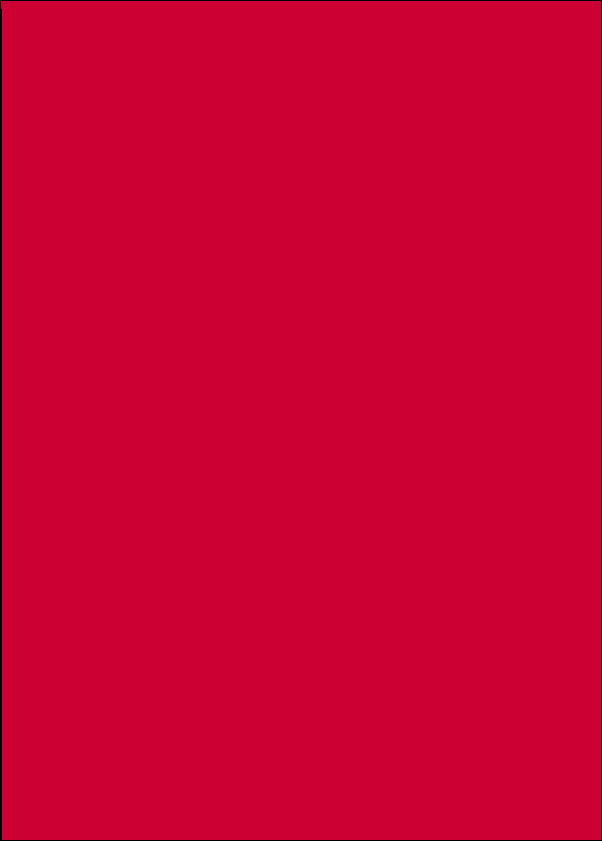
The SMC would represent a voluntary-compliance approach to the oversight of content moderation: participants (social media platforms and all stakeholders) sign up to a mechanism that does not create legal obligations. Its strength and efficiency rely on voluntary action by platforms, whose commitment, when signing up, will be to respect and execute the SMC's decisions and recommendations in good faith. The SMCs would have an advisory and an adjudicatory role. In an advisory capacity, it would provide general guidance on content moderation on the basis of international standards. In that sense, the SMC would be a forum where stakeholders elaborate recommendations. In our experience, when all stakeholders have the opportunity to contribute to the elaboration of the rules (or the interpretation thereof), the awareness of, and abidance with, such rules percolate more naturally through their everyday practices. As such, the collective elaboration of guidelines would contribute to weaving international standards in the understanding and practice of content moderation.

The SMC would also have the power to review individual content moderation decisions made by social media platforms. Such a mechanism would have to be accessible to all, and there needs to be clear and precise rules of procedure on questions such as admissibility conditions, time limits, admissibility of evidence, elements covered by confidentiality, exchange of arguments and views, elements of publicity, and the adoption and publication of decisions. The existence of such a right of appeal would give the SMC more credibility in the eye of the general public.

---

2 See for instance Online Platforms' Moderation of Illegal Content Online, June 2020, available at https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2020)652718; Freedom and Accountability: A Transatlantic Framework for Moderating Speech Online (Philadelphia: Annenberg Public Policy Center, June 2020), available at https://www.annenbergpublicpolicycenter.org/twg; CIGI, Models for Platforms Governance, October 2019, at https://www.cigionline.org/platforms; B. Sander, Freedom of Expression in the Age of Online Platforms, Fordham International Law Journal, Vol. 43:4 2020, 940-1006.

ARTICLE 19 considers that the SMC can play a significant role within a legal framework of co-regulation provided that the respective roles and powers of the overseeing public authority and of the SMC are clearly delineated. In this configuration, a statutory public authority would set general objectives (such as the existence of appropriate processes and measures) while the SMC would provide a space where technical and practical mechanisms and innovations towards these objectives can be discussed with all stakeholders and tested for compliance with international standards on freedom of expression and other fundamental rights. Under monitoring by the statutory authority, the SMC provides some breathing room that facilitates the emergence of a consensus on the appropriate approach towards legal requirements as well as a broader understanding of the complex challenges of content moderation.

ARTICLE 19 believes that the SMC project marks a point of convergence between the goals and interests of human rights groups, civil society activists, users, regulators, policy-makers and social media platforms:

- avoiding the pitfalls of harsh legislative approaches that often come with disproportionate sanctions;
- contributing to restoring trust from users through increased transparency and accountability;
- providing an effective yet adaptable form of regulation that can accommodate the constant evolution of tech platforms;
- and ensuring that effective content moderation complies with the universal principles of international law on freedom of expression and other fundamental rights.

| **Question 26: How might Ofcom best support VSPs to continue to innovate to keep users safe?** | Ofcom could support VSPs to continue to innovate to keep users safe in two main ways:
- supporting a multi-stakeholder dialogue aimed at the definition and exchange of best practices;
- working together with the Competition Markets Authority to ensure that the market remains competitive and open to new entrants. |

An open, fair and competitive market remains one of the main drivers to innovation and to quality improvements in the products/services available for users.

We believe that Ofcom should refrain from creating direct incentives for VSPs to innovate to keep users safe that could lead VSPs to over-remove content, such as setting targets that use metrics like 'how many' and 'how fast'.

In any case, Ofcom should require VSPs to be more transparent about their actions to keep users safe. For full and effective transparency, data should be provided with maximum levels of granularity and a methodology allowing for an effective comparative analysis and evaluation of the content moderation methods applied.

Question 27: How can Ofcom best support businesses to comply with the new requirements?	
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?	ARTICLE 19 welcomes the set of principles set out in paragraph 2.49 and in particular those about safeguarding freedom of expression, adaptability over time, transparency, independence and proportionality. We suggest Ofcom to add an evidence-based approach to this list, which we believe would help the regulator to act in compliance with the principle of proportionality. The evidence-based approach is mentioned in paragraph 2.48 but it does not seem to be provided with the same relevance as the principles in paragraph 2.49.

Moreover, an evidence-based approach, coupled with the principle of proportionality, is necessary for the deployment of the principle of protection and assurance in a way that guarantees freedom of expression. Indeed, we are worried that the principle of protection and assurance might be applied in an unduly broad manner and that the statutory protections for consumers will unduly restrict the latter's free expression right. A careful balancing exercise is needed. A way to do so is to insert, in the statutory protections, mechanisms to check the effectiveness of the measures over time and to remedy unnecessary and disproportionate limitations of free expression on an ongoing basis.