Your response

Question	Your response
Question 1: Do you have any comments on Section 3 of the draft guidance on harmful material and related definitions?	Confidential? – N The definitions provided in section 3.4 of the guidance require further consideration. The inclusion of illegal material under 'relevant harmful material' (3.4, b, ii) is at odds with the Online Safety Bill's categorisation of illegal and legal but harmful material. There is also likely to be significant overlap in content that might "impair the physical, mental or moral development of under-18s" (as defined under restricted material) and relevant harmful material. The guidance states that "Ofcom considers that compliance with the VSP regime will assist services in preparing for compliance with the online harms regime" (2.24). To support compliance and the transition between the regulatory regimes, a clear categorisation of harmful content for children in both VSP guidance and the final Online Safety legislation is desirable. **Recommendation:** 5Rights recommends that illegal content should be defined separately, rather than being included in 'relevant harmful material'. **Recommendation:** 5Rights recommends that 'material that might impair the physical, mental or moral development of under-18s' is defined separately to 'restricted material'- for example as 'content that may be harmful to children.' This provides a useful basis for the Online Safety Bill which will need to define the types of content that are harmful to children. The list of examples of 'material that might impair the physical, mental or moral development of under-18s' (3.11) is welcome as a practical aid for VSPs. However, misinformation and disinformation are currently absent from this list. Ofcom's 2020 News Consumption in the UK report showed that UK children receive most of their information online, predominantly from social media sites that are VSPs[1]. Instagram, one of the most popular sources of news for 16-24-year olds, has recently been found to recommend misinformation about Covid-19, vaccines and

[1] News Consumption in the UK: 2020. Produced by Jigsaw Research, Ofcom. August 2020. Available online.

elections.[2] Children also find it more difficult than adults to identify what is true and what is false.[3] Given the risks posed by this content, misinformation and disinformation are included as a 'content' risk under the 4 Cs framework[4] – a tool that categorises online risks to children into content, contact, conduct or contract (commercial) risks.

Recommendation: 5Rights recommends that misinformation and disinformation is added to the list of 'material that might impair the physical, mental or moral development of under-18s.'

Under the proposed guidance, terms and conditions must state that users uploading restricted content to VSPs are required to bring this to the attention of the VSP provider (2.33 (1) and that 'a person must not upload a video containing relevant harmful material' (2.33 (2)).This puts a huge responsibility on users to read terms and conditions, understand if their content meets the criteria for 'restricted material', correctly classify their content and take the appropriate action. This approach puts a huge amount of responsibility, particularly on those under the age of 18, and would require a radical change in how terms and conditions are presented and understood.[5] Children uploading content may find it more difficult to understand the different categories of material, the actions that are required of them and the consequences of failing to categorise material correctly or notify the provider.

Instead, providers should be required to prevent as far as reasonably possible the uploading of illegal content through the use of moderation and auto-detection technologies at the point of upload, for example PhotoDNA. This would reduce the reliance on users correctly classifying material and notifying the provider and introduce a more robust technical measure to prevent the uploading of illegal material.

Recommendation: 5Rights recommends that the requirements for restricted, illegal and harmful content adequately reflect the impact of this content for children, with the most stringent rules applied to illegal content (i.e do not upload supported with auto-detection

[2] Report by Center for Countering Digital Hate. Malgorithm: How Instagram's Algorithm Publishes Misinformation and Hate to Millions During a Pandemic. 2021, available online.

[3] For more information see Risky by Design by 5Rights Foundation.

[4] Content, contact, conduct and contract – updating the 4Cs of online risk, Children Online: Research and Evidence, February 2021.

[5] Published terms are rarely read by adult users or young people. For example, a Deloitte survey found 97% of 18–34-year-olds agree to terms and conditions before reading them. Available online.

	technologies). This could be included as an addition of a new subsection following 4.23 as follows; New 4.24. In addition to including terms and conditions to the effect that a person must not upload to the service illegal content, providers must utilise available auto-detection technologies, for example PhotoDNA to detect CSAM.
Question 2: Do you have any comments on the draft guidance about measures which relate to terms and conditions, including how they can be implemented?	Confidential? – N We support the five principles to support implementation (effective, easy to use, transparent, fair and evolving). We particularly welcome the provisions set out in 4.29 for child-friendly explanations.[6] However, the guidance should stress that clear presentation of terms and conditions alone does not equate to effective implementation. As well as clear presentation, terms and conditions must be upheld and enforced. Interface-level nudges like tools that filter offensive words and monitoring tools to detect harmful keywords or phrases can help services to enforce their terms and conditions and support users. For example, R;pple Suicide Prevention is an online monitoring tool currently in development that redirects users searching for harmful keyword or phrases to different forms of mental health support.[7] The R;pple tool acts as an interception, guiding users away from self-harm or suicide content towards mental health resources. This will provide a much stronger intervention than is currently offered by search engines and demonstrates the kind of mechanisms available to protect users from harmful content. **Recommendation:** 5Rights recommends that in addition to clear presentation, published terms are supported by: • Swift and easy access to expert advice • Swift and easy access to redress • Preventing automated recommendation of harmful material • Disabling any design features that connect children with unknown adult users

[6] 5Rights' forthcoming publication, 'Tick to Agree: Age Appropriate Presentation of Publish Terms' sets out how terms and conditions can be better presented to children and young people to maximise comprehension and ensure meaningful consent.

[7] How the R;pple tool works, R;pple Suicide Prevention, 2021.

	• Restricting design features that extend use, including autoplay or timed notifications, particularly at night • Positive nudges to encourage 'time off' • Ensuring informed consent (for example providing transparency about the risks associated with the nature and features of product or service). Consent needs to be meaningfully obtained from children, and where necessary, parents.
Question 3: Regarding terms and conditions which prohibit relevant harmful material, do you have any comments on Ofcom's view that effective protection of users is unlikely to be achieved without having this measure in place and it being implemented effectively?	Confidential? – N Yes, 5Rights agrees with this view. However, effective protection will not be possible without minimum standards for terms and conditions, supported by regulatory oversight and enforcement. We welcome the inclusion of moderation processes (and their effectiveness) in the recommendation to conduct risk assessments (4.45), particularly 4.42 which sets out Ofcom's expectation that moderation should be subject to quality assurance and "the accuracy of any machine-learning moderation techniques to be checked using human quality assurance processes."
Question 4: Do you have any comments on Ofcom's view that, where providers have terms and conditions requiring uploaders to notify them if a video contains restricted material, additional steps will need be taken in response to this notification to achieve effective protection of under-18s, such as applying a rating or restricting access?	Confidential? – N Yes, 5Rights agrees additional steps will need to be taken. The guidance gives examples of such steps as tagging or rating mechanisms (4.79-4.86) and access control measures including age assurance and parental controls (see 4.22 and 4.89). Though these are welcome additional steps, the guidance should reinforce the need for providers to take a holistic approach to the protection of under 18s and consider system design and functionalities that create risks. This point is made very clearly by Ofcom in the consultation document accompanying the guidance in sections 2.12 and 3.31, explaining VSPs will need to "consider how to mitigate the risks that may lead to harms." This would also align with the draft Online Safety Bill, in which one of the 'online safety objectives' includes "designing and assessing the service with a view to protecting users from harm, including algorithms, functionalities, and other features relating to the operation of the service." 5Rights' project, Risky by Design[8] illustrates how such design features can

[8] For more information see Risky by Design by 5Rights Foundation.

	pose risks to children and why services must mitigate these risks to protect users from harm. **Recommendation:** 5Rights recommends the guidance include a requirement for providers to assess how aspects of design, for example, recommendation systems and autoplay, contribute to the risks under-18s are exposed to on the service, and then take steps to mitigate these risks. **Recommendation:** 5Rights recommends that the language in relation to additional steps for effectively protecting under-18s is strengthened, for example, in 4.45 as follows; 'Ofcom ~~may~~ will use this information to support its compliance monitoring (Section 7)'.
Question 5: Do you have any comments on the draft guidance about reporting or flagging mechanisms, including on Ofcom's view that reports and flagging mechanisms are central to protecting users?	Confidential? – N It should not be expected or assumed that a child will be able to identify or report content or conduct which are against a service's community guidelines. Children may find some of the rules set out in community guidelines confusing or struggle to distinguish between what is illegal and what is legal but prohibited by a service.[9] They may not know if they themselves have breached a service's terms or what to do when something goes wrong, or how a service will respond when they have a problem that needs attention. They may be hesitant to report problems if they are concerned that they will get into trouble. While swift, effective reporting is an important provision for children, it is not the 'central' mechanism for protecting users. Relying on user reporting requires a child to understand the harm and their rights to be treated differently. This is simply not the reality for many children, who may feel shame, who may not understand what is happening, who might be scared their device will be taken away, or who may not trust that the system will take care of them. It is welcome that the guidance encourages providers with a high number of users under the age of 18 to consider the needs of this group when designing or reviewing reporting/flagging systems (4.60), but this should be a requirement, not something providers only need to 'consider'. **Recommendation:** 5Rights recommends that the guidance includes a requirement for providers to implement the highest protection measures (not only reporting or flagging mechanisms) on services likely to be accessed by under-18's.

[9] Online abuse: teenagers might not report it because they often don't see it as a problem LSE blog by Powell-Jones. May 7th 2019. Available: online.

Question 6: Do you have any comments on the draft guidance about systems for viewers to rate harmful material, or on other tagging or rating mechanisms?	Confidential? – N Section 4.76 explains that "platforms might allow viewers to challenge and ultimately change ratings. For example, if enough viewers believe that a different rating should be applied, this might be amended without intervention from the platform being required. Platforms may also use these rating systems to test or improve the algorithms or other mechanisms which recommend content to users." As noted in 4.77, there are significant risks associated with this approach, including "accuracy and gaming the system" which are valid concerns given the risks to children associated with viral video content. This year, a 10-year-old girl from Italy died in January[10] and a 12-year-old boy died in April[11] after reportedly participating in a 'Blackout challenge' that gained popularity on TikTok. While 4.78 notes that "crowdsourcing" is not expected to provide adequate protection when used in isolation from other measures, given the risks of this approach, the option for users to challenge and change ratings before intervention from the platform should be removed for services accessed by children. **Recommendation:** 5Rights recommends that the guidance explicitly states in a new subsection following 4.28 that 'for services likely to be accessed by under-18s, VSPs should not allow viewers to challenge or change the rating of harmful material before being reviewed approved by moderators.'
Question 7: Do you have any comments on the draft guidance about age assurance and age verification, including Ofcom's interpretation of the VSP Framework that VSPs containing pornographic material and material unsuitable for classification must have robust age verification in place?	Confidential? – N 5Rights agrees that VPSs containing pornographic material and material unsuitable for classification must have in place "a robust access control system that verifies age and prevents under-18s from accessing such material." We are also encouraged that the guidance sets out a proportionate approach to age assurance based on an assessment of risk, with provisions that the most harmful restricted material – not only pornography, but any material that might impair the physical, mental or moral development of under 18s – requires a high bar of age assurance. The considerations for effective age assurance (4.108) are both comprehensive and practical, recognising that integrated age assurance mechanisms that minimise disruption to the user experience are more likely to be

[10] 10-year-old girl dies after TikTok blackout challenge, Brussels Times, January 2021.
[11] Boy, 12, dies after doing TikTok Blackout challenge. April 2021. The Independent.

adopted, and that for smaller services, a third-party solution may be the most practical way to achieve the required level of confidence in the age of their users. We also welcome consideration of potential exclusionary risks to children and acknowledgement of the limitations of certain approaches, such as self-declaration, that do not account for cases where an under-18 might provide false information to bypass age assurance measures.

However, the requirements for higher risk services must ensure that VSPs that carry the greatest risks to children have in place the most robust forms of age verification.

Recommendation: 5Rights recommends redrafting 4.108 to read;

a) It is important to assess, in a privacy preserving way, who is using the service. Higher risk services ~~should make greater efforts~~ are required to have the highest levels of confidence in ~~understand~~ the age of their users, so that appropriate access control measures can be established and operated.

Recommendation: Additionally, 5Rights recommends the addition of the following to the list of age verification approaches Ofcom would **not** consider to be appropriate protection measures for material of a pornographic nature:

Cross account authentication

Cross account authentication (using an existing account to gain access to a new product or service) can provide convenience for users by removing the need to prove their age every time they access a service, but provides varying levels of assurance to providers in the age of their users. The level of assurance is determined by the method used by the original authenticating provider and is not therefore an appropriate method of age assurance for high-risk services or material.

And

Profiling (AI)

Profiling, or the use of AI for age assurance, creates a significant tension between data processing and a child's right to privacy. It does not provide a high level of assurance in the age of a user if the quality of the data is poor or the dataset contains errors or omissions. Profiling is also likely to result in the collection of data beyond that which is needed for age assurance, and there is a risk that the data derived from profiling will be shared with third parties and used in ways that has a detrimental impact on children.

Question 8: Do you have any views on the practicalities or costs relating to the implementation of robust age verification systems to prevent under-18s from accessing pornographic material and material unsuitable for classification? Please provide evidence to support your answer wherever possible.	Confidential? – N The technology for age assurance is already available, but its widespread adoption has been hampered by a lack of public trust in the practices of data-hungry tech companies, and the absence of minimum standards set out in regulation.[12] The issue is less one of practicality or cost but of trust, oversight and enforcement. Ultimately, the cost or practicalities of one solution over another should never come before the requirement to provide the most effective and appropriate form of age assurance that is proportionate to risk and purpose. **Recommendation:** 5Rights recommends redrafting 4.90 to read; In determining an approach to obtaining appropriate assurances as to the age of potential viewers, we ~~encourage~~ require VSP providers to conduct a risk assessment of their platform, and select an approach (or approaches) that is proportionate to risk and purpose. ~~having regard to the practicable and proportionate criteria~~
Question 9: Do you have any comments on the draft guidance about parental control systems?	Confidential? – N We are concerned that the VSP guidance gives age assurance and parental controls equal weighting. VSPs should not be able to defer responsibility for the safety of children to parents and carers through parental controls, instead of making their services safer by design and default. Parental controls are not a substitute for good design that prioritises user safety. They tend to concentrate on adult anxieties such as screen time, whilst overlooking less obvious risks to children, such as contact and conduct risks when livestreaming. In this way, parental controls may offer false security to parents while children continue to be exposed to risks due to poor service design. Parental controls do not always account for mixed ages within the same family group. If controls are designed primarily for a younger age groups, they can limit the experience of the digital world for older children or result in access being overly and unfairly restricted. Similarly, paired accounts between a parents or carer and a child, common on streaming platforms, may be suitable for younger children but not for older children who wish to

[12] See: Age Assurance (minimum standards) Bill, House of Lords, May 2021

	have a degree of autonomy and privacy from their parents. The linking of parent/carer and child accounts also raises concerns about additional profiling and data collection, including a service's ability to link a child's network to an adult's network. While we are encouraged by the provision in 4.113d that states "under-18s should be able to enjoy age-appropriate content and activity on a service without undue parental interference," we remain concerned that an over reliance on parental controls will preclude the development of services that are safer-by-design, and other more equitable protective measures, such as age assurance. **Recommendation:** 5Rights recommends that under the 'Parental Controls' heading, the text in the grey box is revised to read; "~~Provide for~~ Consider offering parental control systems in relation to restricted material as a supplement to age assurance mechanisms. **Recommendation:** 5Rights recommends that account holder confirmation (paired accounts) and parental controls are included as alternatives to age assurance, only in limited cases where age assurance is not possible or not appropriate. As currently drafted, the guidance suggests that all VSPs with child users should consider offering parental controls on their services. We do not support this approach as in the case of services directed at children, this would risk undue interference from parents. **Recommendation:** 5Rights recommends 4.109 is redrafted to read; "Providers of mixed audience services ~~who offer services to under-18s should strongly~~ may consider having some form of parental control feature to support their overall protection measures for under-18s."
Question 10: Do you have any comments on the draft guidance about the measure regarding complaints processes or on the regulatory requirement to provide for an impartial dispute resolution procedure?	Confidential? – N VSPs must be required to establish and operate a complaints procedure that is transparent, easy to use and effective, and which does not affect the ability of a person to bring a claim in civil proceedings. 5Rights research shows that often young people do not know if they have breached a service's terms, or what to do when they feel something is wrong: *"Nothing happens on social media, no one gets into trouble. So many people say bad stuff."*

	"If we have a problem, encourage us to say so and tell us exactly what you are going to do about it… and how quickly." *– Young person, UK* VSPs must set out clearly the kinds of content, activity and behaviour that is and is not permissible on their service. Flagging and reporting tools must be clearly signposted and young people should be offered specialist support and advice to help them understand the nature of their complaint, their rights, their choices and the help that is available. Users should be told when they can expect a response or action and how that response will be communicated after a report is made. The response timing should be appropriate to the seriousness of the report being made. For example, reports relating to young people's safety should take priority over reports relating to copyright concerns.[13] A triage system is also useful to streamline priority responses and determine which complaints require human interaction/evaluation. In considering complaints from children, VSPs need to take into account the diversity of children's experiences: what is acceptable to one child may be deeply upsetting to another. Children's complaints should be prioritised and children who appear to be in distress, even for minor infringements, should be responded to quickly and offered appropriate support. **Recommendation:** 5Rights recommends the addition of the following text under 'Complaints process' (4.114 - 4.116) Complaints relating to the protection and safety of under 18s must be prioritised, and VSP's must have the capacity and systems in place to respond immediately to serious incidents that might involve child protection or safeguarding concerns, or criminal activity.

[13] YouTube have a wide variety of tools for users to file copyright complaints including a 'Copyright Match Tool' and a digital fingerprinting system to scan content. If a copyright owner submits a valid DMCA (Digital Millennium Copyright Act) complaint, YouTube will take down the content and apply a copyright strike to the user who uploaded it.

A user's account is terminated, along with any associated channels if they receive three copyright strikes in 90 days. Compare this to Youtube's response to harassment complaints where a user is sent an email notification if their content is removed on the basis it breaches community guidelines. If it is their first violation, a user will only receive a warning with no penalty, and a strike is only issued against their channel for future violations. See: https://www.youtube.com/howyoutubeworks/policies/copyright/#making-claims

Question 11: Do you have any comments on the draft guidance about media literacy tools andinformation?	Confidential? – N Generally, this section is strong and it was particularly encouraging to see the inclusion of media literacy relating to the recommendation of content, recognising different types of content (advertising or 'paid-for' content) and the business models of platforms (use of data to inform targeted advertising) - 4.152 b i, ii, iii. Additionally, we welcome the recommendations to 'bake in' media literacy throughout the user journey and at the level of design, as set out in section 4.153 b. **Recommendation:** 5Rights recommends that 4.159 explicitly refer to the evolving capacities of children and recognising their developmental needs. This is captured in section 2.21 of the consultation document that accompanies the draft guidance, but would support VSPs in their efforts to promote media literacy if included as follows; 4.159 For example, the COM-B model can help with understanding how the physical or psychological capabilities of a user might impact their propensity to engage with reporting and flagging mechanisms, whether the physical environment (opportunity) encourages or discourages users to engage with terms and conditions, or how a user's motivation may deter them from seeking media literacy guidance or information. VSPs must also consider the developmental stages and evolving capacity of under 18s to ensure media literacy tools are pitched at an appropriate level to meet their needs. **Recommendation:** 5Rights recommends that children (and parents of younger children) are informed about the full range of risks presented by the service. For example, 4.152 could be expanded; iii) Awareness of the potential benefits and possible risks of using the VSP. For example, information may be provided about persuasive design features that are being used on the VSP.
Question 12: Do you have any comments on the with the draft guidance provided about the practicable and proportionate criteria VSP providers must have regard to when determining which measures are appropriate to take to protect users from harm?	Confidential? – N The guidance (5.1) names 'the size and nature of the video-sharing platform service' as first in the list of consideration for determining which measures are appropriate. Reference to the nature of a service is appropriate, however a small service is not necessarily a safe service.[14] Smaller providers need greater support to comply with regulation, not permission to harm. This is the norm in

[14] Small services that carry risk such as the video-sharing platform Clapper, which has under 100,000 downloads on the Google Play store. Despite a minimum user age of 17, the service's weak age assurance means a child can log into Clapper via their Google account, even if they are underage. The service is known to harbour misinformation and its terms of service explicitly state that it "cannot ensure the prompt removal of objectionable material as it is transmitted or after it has been posted."

	other industries, for example health and safety regulations at work, where measures are determined by the nature and size of the risk, not the nature and size of the employer. **Recommendation:** 5Rights recommends editing 5.1 as follows 'a) the ~~size and~~ nature and risks inherent to the video-sharing platform service;' It is very welcome that 5.30-5.32 of the guidance explains that VSPs should take into account "the rights and legitimate interests of users." **Recommendation:** 5Rights recommends adding a section following 5.32 as follows; New 5.33 Children's rights, as defined by the United Nations Convention on the Rights of the Child (UNCRC) and set out in the recently adopted General Comment No.25[15] in relation to children's rights in the digital world, must be upheld by providers in accordance with the UK's obligations as a signatory to the UNCRC. The best interests of the child should be prioritised when designing and operating systems and processes.
Question 13: Do you have any comments on the draft guidance about assessing and managing risk?	Confidential? – N We welcome the acknowledgement of risks associated with design as referenced in 3.11e, including "AI and algorithmic manipulation; profiling and persuasive design including nudging and targeting leading to a detrimental impact on under-18's". In the proposed guidance, 'assessing and managing risk' is included as an additional measure (6.23). It is welcome to see Ofcom go beyond the VSP framework by including this in the proposed guidance, however, risk assessments should be mandated for services likely to be accessed by children, as they are in the Online Safety Bill. Including a requirement to conduct child risk assessments would also support Ofcom's own view that 'in particular, where under-18s form a significant proportion of the user base, they would be a key factor for consideration in the risk profile because they are a clear category of persons to be protected under the VSP Framework (see Section 3).' **Recommendation:** 5Rights recommends the guidance requires VSPs likely to be accessed by children to conduct

[15] General Comment on children's rights in relation to the digital environment, OHCHR, March 2021.

	child risk assessments instead of more generally 'assessing and managing risk'.[16] **Recommendation:** 5Rights recommends the guidance state VSPs are **required**, rather than encouraged, to anticipate the risks posed by the design of their service,. This this could be reflected in, for example; 2.37c) ii) Section 6 of this Guidance outlines ~~encourages the use of additional~~ steps to help protect users. These include considering the ~~practicable and~~ proportionate criteria and decisions about protection measures as part of a risk management process. This process should involve identifying potential harms on a platform (including both intended and unintended consequences of recommendation systems and automated-decision making; persuasive design including nudges and targeting that has a detrimental impact on under-18s); documenting decisions about the measures in place to mitigate those potential harms; and measuring the effectiveness of those measures.
Question 14: Do you have any comments on the impact assessment in Annex 1, including the potential impacts to VSPs outlined in tables 1 and 2, and any of the potential costs incurred (including any we have not identified)?	Confidential? – N No comments to add.
Question 15: Do you have any comments on our provisional assessment that the potential costs for providers are proportionate to achieve the regulatory requirements of the regime?	Confidential? – N 5Rights agree that the costs are proportionate.
Question 16: Do you have any comments on any other part of the draft guidance?	Confidential? – N We have some concerns around the status of the guidance and how Ofcom will assess compliance with the VSP regulation that the recommendations are designed to support. Most of the provisions contained within the guidance are recommendations, and there are very few instances where it is explicitly stated a provider **must** take a certain course of action. As stated in 4.6 "we have provided guidance on what platforms "should do" or "should consider" when

[16] See section 7, 3) and 4) of the Draft Online Safety Bill. Presented to Parliament by the Minister of State for Digital and Culture by Command of Her Majesty, May 2021.

	implementing measures in a way that achieves the requirement to protect users. These are not prescriptive requirements but intended as helpful suggestions to aid understanding of how compliance could be achieved. In some instances, there may be other ways to implement a measure to achieve the same requirement. Where we think effective protection of users is unlikely to be achieved without a specific approach, we say so." While there are sections that indicate where a measure is unlikely to meet the requirements of the VSP regulation, the guidance does not go far enough to mandate certain actions from providers to implement the required level of protection. This will not only create a challenge for Ofcom in enforcing the regulation but make compliance with the regulation more difficult for providers who may not be certain of the appropriate measures to take. The VSPs to whom this regulation applies will soon be subject to the duty of care brought in under the Online Safety Bill. To support those providers, this guidance should reflect the duties introduced in the Bill, such as the requirement to conduct child risk assessments, and vice versa, the duties in the Bill should reflect certain additional requirements for VSPs set out in this regulation. In particular, the requirements for VSPs to establish and operate age verification to prevent under 18s accessing pornographic and restricted material should be replicated under the new online safety regime.