NSPCC Response to Ofcom's consultation on Video-Sharing Platform Guidance

May 2021

The NSPCC welcomes the chance to respond to Ofcom's consultation on its draft Guidance for video sharing platforms (VSPs). The introduction of regulatory requirements on video sharing services, resulting from the implementation of the Audio Visual Media Services Directive (AVMSD), presents an important opportunity to secure robust protections for children using high-risk video services and design features.

The development of Ofcom's regulatory scheme will inevitably attract a high level of interest, given the VSP regime will serve as a set of interim arrangements ahead of the introduction of the provisions set out in the draft Online Safety Bill.

Ofcom's stated intention to build on VSP regulation to inform its approach to online harms regulation means that its extent and ambition will not only determine initial protections to under 18s using VSPs, but will send a wider message to platforms likely to be in scope of the online harms regulation about the regulator's approach and risk appetite.

The NSPCC looks forward to working closely with Ofcom on the development of its regulatory approach.

Harmful material

Question 3: do you have any comments on section 3 of the draft guidance on harmful material and related definitions?

Section 3 proposes harmful material should be considered in two overarching categories: 'relevant harmful' material that includes illegal content; and 'restricted' material which refers to legal but potentially harmful content. Such content includes age-inappropriate material for children and young people such as pornography, and material that might otherwise impair the physical, mental or moral development of children.

Although we welcome the proposed definitions, we have significant concerns about how the approach reflects the dimensions of the child abuse threat.

In particular, we note the draft Guidance adopts an overly narrow definition of child sexual abuse material. For the purposes of defining 'relevant harmful material', the Guidance states that it considers 'the offences […] most relevant for VSP providers to be related to the distribution, dissemination or transmission of child pornography'.

As a result of adopting a definition that focuses so tightly on the distribution and dissemination of child abuse imagery, the proposed definitions for 'relevant harmful' material fail to capture the circumstances in which child abuse images are being actively *produced* on VSPs, and to therefore appropriately reflect the broad dynamics of illegal behaviour.

Self-generated content is an increasingly significant driver of child abuse images: of the 153,369 webpages actioned by the Internet Watch Foundation (IWF) during 2020, almost half (68,000 or 44%) were assessed as containing self-generated imagery. This is a 16% increase on the total number of reports actioned in 2019.[1]

Self-generated images can be produced as a result of clearly illegal activities, including grooming. Images can also be produced by children themselves, either consensually or non-consensually, and in either circumstance can go on to be consumed as child abuse images. Self-generated imagery can include photographs and videos, as well as livestreamed content which is produced on VSP services but which can be captured as a standalone image or video.

Under UK law, the production of self-generated imagery sits within a complex legal picture, although in the strictest of terms, if a child takes an image of themselves that meets the criminal threshold to be considered a child abuse image, they have committed an offence, regardless of context.

The NSPCC is therefore concerned that the proposed definition of 'relevant harmful' material fails to appropriately reflect the dynamics of the production and distribution of online child abuse, when it is increasingly difficult to separate out the circumstances in which images are produced and shared.

Furthermore, the proposed definition sits uncomfortably with the current UK legal framework for child abuse offences. It is highly questionable that clearly illegal behaviour such as grooming should be considered only as 'restricted material.' Similarly, the classification of all forms of 'sexting, naked selfies and nudes' within the same categorisation fails to consider and appropriately specify forms of behaviour which more appropriately should be considered 'relevant harmful', or understand the full complexity associated with it.

We strongly encourage Ofcom to review its Guidance to better align against the UK and EU legal frameworks, and to recognise and respond to the significant changes in how child abuse material is produced and shared online in the decade since the CSEA Directive was passed.

We would also note other relevant legislative mechanisms and decisions in this regard, including the provisions of Article 8.[2]

Under the VSP Framework, providers will be required to protect all users from relevant harmful material, and children from restricted content. Although this means VSPs should be required to address child abuse material regardless of its classification, in practice many platforms may be likely to adopt a differential approach to compliance.

Beyond the immediate issues around legal alignment, and our developing understanding of the child abuse threat, it is vital that the Guidance encourages VSPs to consider child abuse risks as an overarching and highly related set of issues. However, the differentiated categorisation of child abuse risks may detract from this objective.

---

[1] Internet Watch Foundation (2021) 'Grave threat' to children from predatory internet groomers as online child sexual abuse material soars to record levels'. Cambridge: IWF.

[2] Under article 8 of the European Convention on Human Rights, states have a positive obligation to protect the physical and psychological integrity of an individual from other persons. This applies particularly to the well-being of vulnerable persons, and in order to protect their right to a private life, includes the protection of the child from physical and mental harm. States are under a duty to take action to address those risks (including whether state is not primary violator). See for example KU vs Finland (2015), or O'Keeffe vs Ireland (2014) in the European Court of Human Rights

For example, in the event that companies sought to prioritise compliance against relevant harmful material, because of its close(r) alignment with illegal forms of content, this could result in a less developed response to the highly pressing and related issues of online grooming and the production of new child abuse content.

The NSPCC considers these the detection and disruption of grooming and the production of self-generated child sexual abuse imagery should be a priority focus for Ofcom's scheme, with such harms being reclassified in the Guidance as being of a 'relevant harmful' status.

Terms and conditions

Questions 2 and 3: do you have comments on the draft guidance about measures which relate to the conditions?

Regarding relevant harmful material, do you agree that effective protection of users is unlikely to be achieved without having this measure in place and implemented effectively?

We strongly support the requirements for VSPs to introduce measures that prevent children's exposure to relevant harmful restricted material, and to enact them in such a way is to carry out the relevant purpose. These should be subject to an effective and transparent risk management process, as set out in section 6.

We endorse Ofcom's assessment that having and enforcing clear terms and conditions relating to relevant harmful material is fundamental to the VSP regime. It is highly unlikely that effective protection of users can be achieved without having this measure in place and being implemented effectively.

We are concerned that some VSPs likely to fall within UK jurisdiction do not have adequate arrangements in place to uphold and enforce their terms and conditions. For example, OnlyFans appears to have highly ineffective child abuse detection and disruption mechanisms in place.[3]

In 2020, the most recent year for which data is available, OnlyFans made zero reports to the National Centre for Missing and Exploited Children (NCMEC)[4]. Our understanding is that the platform has opted not to take the full range of available industry 'hashlists', and at present time, is not a member of the Internet Watch Foundation (IWF.)[5]

Given the particularly egregious impacts of child abuse, and the circumstances relating to how child abuse material is produced and distributed on VSPs, we recommend the Guidance states more explicitly what moderation and enforcement expectations will be considered necessary for a VSP to demonstrate its compliance.

As a minimum, we recommend the Guidance clearly states that VSPs will permanently remove and delete accounts that share child sexual abuse material, and will report any illegal activity relating to child abuse to law enforcement.

---

[3] Recent BBC investigations have found that children have been exposed to a range of risks on the OnlyFans platform, including being able to produce and post self-generated content, and as a result being exposed to additional risks including grooming and exploitation by other users.
[4] NCMEC (2020) Reports by ESPs.
[5] Based on discussions with the Internet Watch Foundation

Companies should be required to demonstrate they have proactive and child-centred processes in place to proactively detect child abuse. VSPs should be able to demonstrate in their terms and conditions that children can request the takedown of images which are likely to cause them distress, and that these processes are easy-to-use, transparent and result in swift action.

At present, we are aware that some VSPs may not takedown images even where the criminal threshold is met[6] and where there the subject of a child abuse image has repeatedly reported it to the platform involved. Some VSPs have wholly ineffective reporting processes for child abuse, which can be highly challenging to navigate and successfully use.

Question 4: do you have any comments on the view that, where providers have terms and conditions which require uploaders to notify them that the video contains restricted material, additional steps will be needed to achieve effective protection of under 18s?

We agree with the assessment that, where providers choose to take this measure, it is unlikely they will achieve effective protection for under 18s without taking additional measures to either notify viewers where video contains restricted material, or the case of commercial or user generated pornography, to restrict access to it by under 18s.

It is reasonable to expect that VSPs should introduce this measure in conjunction with effective age assurance mechanisms, meaning that services can ensure content being accessed by or actively recommended to children across different age bands is age appropriate.

Although this approach may offer some additional protections to children and young people, if the measures taken simply require an age gate which requires children to click through before accessing content, this is unlikely to offer substantive protections in practice.

As part of their quality assurance processes, VSPs that adopt this measure should audit the extent to which users comply with the requirements to categorise restricted material. VSPs set out what, if any, action is taken against users who persistently fail to label or tag their posts correctly.

VSPs should only look to introduce or rely upon this measure only where it is cost-prohibitive or technically infeasible to deliver other solutions that could more accurately and reliably identify restricted content and prevent children's exposure to it, for example high quality AI classifiers.

Reporting and flagging mechanisms

Question 5: do you have comments about reporting or flagging mechanisms, including on Ofcom's view the report and flagging mechanisms are central to protecting users?

We strongly support the VSP regime requiring platforms to develop more easy-to-use and visible complaints and reporting systems for harmful content. NSPCC research finds that a significant proportion of young people aged between 13 and 17 do not know how to make reports on video

[6] Research undertaken by the Canadian Centre for Child Protection shows that some platforms routinely refuse to take down child abuse images where the child displays any physical signs of sexual maturation, meaning that images of children aged nine or 10 may not be immediately removed Canadian Centre for Child Protection (2019) How we are failing children: changing the paradigm. Winnipeg: CCCP

sharing platforms. Less than three-fifths (58%) of young people using Snapchat, and 57% of those using TikTok and Twitch, know how to make a report.[7]

However, there are also significant barriers associated with children feeling suitably confident and willing to use reporting functions. Although research finds that a significant number of children have been exposed to inappropriate content and behaviour[8], only a very small portion of children aged 13-17 recall having made a report, including 3% of young people using TikTok, 4% on Snapchat and 5% using Twitch.[9]

We agree with Ofcom that reporting mechanisms are likely to be fundamental to the protection of users, and that it is unlikely that effective protection of children can be achieved without better reporting mechanisms being in place. It is therefore essential that the regime requires companies to address the visibility and usability of reporting services, but also the barriers to securing high(er) levels of user engagement including among children and people.

To that effect, we welcome the proposed Guidance that VSPs must ensure their reporting and flagging tools are visible, accessible and easy-to-use.

Research undertaken by the Canadian Centre for Child Protection finds a number of significant barriers in the reporting of child abuse material. For example[10]:

- reporting structures can create strong disincentives for users to report illegal content (such as requirements to provide personal contact information);
- many platforms require users to navigate unnecessarily complex and lengthy user journeys before they can make a report;
- some sites may require users to create or log into an account on the platform before they can report publicly visible content;
- poor user design on many VSPs means there may be additional challenges in being able to make a report, including inconsistent navigation between desktop and mobile versions; or the inability to report combinations of specific users, user profiles, or pieces of content.

Although we welcome the requirements set out in the draft Guidance, it seems likely that additional measures will be required to ensure VSPs improve their reporting processes. For example, Ofcom might wish to consider engaging with other regulators on identifying sectoral best practice. This could include working with UX and digital delivery teams to produce replicable design patterns for user reporting mechanisms.

We welcome the requirement in the Guidance for appropriately timely responses to reports and flags, and Ofcom's expectation that platforms should prioritise the most harmful categories of content, including child abuse.

However, the Guidance does not go far enough in setting out clear expectations to VSPs about how and in what form users should be informed or updated about decisions taken in relation to their report.

---

[7] NSPCC research based on interviews with 2,000 young people aged 13-17, on file

[8] Ofcom's own research illustrates that 81% of 12 to 15-year-olds have had a potentially harmful experience online in the previous 12 months. Ofcom (2020) Internet users' experience potential online harms: summary of survey research. London: Ofcom

[9] NSPCC research on file

[10] Canadian Centre for Child Protection (2020) Reviewing the child sexual abuse material reporting functions on popular platforms. Winnipeg: CCCP

Although it is welcome that platforms will be expected to have systems to provide information once a final decision is made, guidance is also required regarding the quality of these notifications. Current notifications are often exceptionally poor, and provide little if any information about why report has been rejected or held, and how a user might appeal a decision (if this is even possible.)

Improved decision notifications are likely to be a key driver of improved user engagement and confidence, including among children and young people.

Age assurance systems

Questions 7 and 8: do you have comments on the draft guidance about age assurance and age verification, including Ofcom's interpretation of the VSP framework that services containing pornographic material and material unsuitable classification must have robust age verification in place?

Age verification for pornographic material

The NSPCC supports Ofcom's interpretation of the VSP Framework that if a VSP has restricted material on its service that is of a pornographic nature, providers should have a robust access control system in place that verifies a user's age and prevent under 18s from accessing such material.

This reflects the clear requirements of the AVMSD that VSPs should introduce the strictest access control measures in relation to material that has the most potential to harm physical, mental or moral development of children.

We agree with the proposed approach that material should reasonably be considered pornographic if its primary purpose is sexual arousal or stimulation. We would welcome the Guidance explicitly stating this applies to both commercial pornography and user generated material.

Ofcom should set out in its Guidance the threshold at which access control measures should be required on VSPs that do not necessarily specialise in adult content, but feature large volumes of it. Some services, for example Twitter, host significant amounts of age inappropriate user generated sexual content, which can be actively recommended and amplified by the site's algorithms.

At present, the Guidance sets out that VSPs must introduce access control measures if they have a 'high prevalence' of pornographic material. Ofcom should clarify what is likely to constitute high prevalence, given the keen interest in the approach taken by the VSP framework in the context of the forthcoming online harms regulation; and to provide regulatory certainty to VSPs likely to be in scope of the regime.

Age assurance processes for restricted material

If VSPs are to effectively prevent children and young people accessing restricted material, and content that may be age appropriate for specific age groups, it is essential they introduce age assurance technologies.

Other measures proposed in the Guidance, including user flagging of restricted or age-appropriate content, can support a more age appropriate experience of children and young people, but it is extremely unlikely that these measures can provide adequate protection from restricted material unless they are accompanied by age assurance mechanisms.

Although the range of technological approaches to deliver age assurance is still developing, it is reasonable to expect that VSPs will already be at an advanced stage of introducing age assurance measures to comply with the ICO Children's Code.

This work is still ongoing, making it appreciably challenging to set out more defined regulatory parameters at this stage. However, it will be important that future iterations of the Guidance provide clearer regulatory expectations about the outcomes expected from age assurance technologies.

Future iterations should include the certainty thresholds that platforms are expected to achieve, and offer guidance on how VSPs are expected to balance trade-offs on overall levels of certainty with the potential for exclusionary effects.

Without clearer guidance, we have substantive concerns about whether companies will feel they have adequate regulatory certainty to invest in more substantive age assurance mechanisms.

VSPs may have legitimate concerns about whether any age assurance mechanisms they intend to put in place can provide a level of certainty about their child users that corresponds with the regulatory expectations likely to be required of them.

Risk management

Question 13: do you have any comments on the draft guidance about assessing and managing risk?

We strongly support VSPs being encouraged to adopt a 'safety first' approach to the design and delivery of their services. Ofcom is right to highlight the importance of a culture of safety, in which the potential risks to users are understood across businesses, and risk mitigations begin at the design stage.

User safety must be hardwired into VSP decision-making, but our experience is that user safety is often siloed within Trust and Safety teams. Ofcom may therefore want to explore how it can use the VSP framework to incentivise a cultural shift in how platforms understand and respond to risks.

For example, the Guidance could explicitly highlight that having clear accountability for user safety at senior levels is an important enabler for VSP compliance, and that it intends to assess VSP governance processes as part of its ongoing monitoring arrangements.

We support VSP providers adopting a clearly defined risk management process. Although this is not a requirement under the VSP regime, it is difficult to see how VSPs could otherwise ensure they have a sufficiently dynamic, targeted risk mitigation regime in place.

Platforms could be encouraged to share their risk assessment with the regulator and relevant expert bodies, recognising this as an important step to build transparency.

We welcome Ofcom's guidance that VSPs should develop a series of metrics to enable them to measure the effectiveness of their regulatory response. Although the proposed list is generally sound, VSPs should be encouraged to place greater emphasis on the prevalence of harmful material.

Metrics should provide an assessment of: how many users see relevant harmful content before it is actioned; the interplay between exposure to harmful content and content recommendation systems; and the differential exposure to relevant harmful and restricted content, including among children and young people.