Written evidence submitted by NSPCC

Response to the DCMS Sub-Committee on Online Harms and Disinformation Inquiry into the draft Online Safety Bill

September 2021

The NSPCC welcomes the Sub-Committee's continued scrutiny of the draft Online Safety Bill, and is pleased to submit evidence to this inquiry.

The draft Bill has the potential to deliver a robust but proportionate regulatory regime, through the adoption of a systemic framework that requires platforms to proactively identify and mitigate potential risks to children.

However, as it stands, we have significant concerns about whether the Bill displays the necessary ambition to tackle preventable child abuse, and to meet the Government's ambition that the UK is the safest place in the world for a child go online.

It is vital that regulation can effectively respond to the growing scale and complexity of technology facilitated abuse. Recent NSPCC data shows that online grooming offences in 2020/21 reached a record high, with sexual communication with a child offences in England and Wales increasing by almost 70 per cent in three years.[1]

This response sets out where we identify significant concerns with the draft legislation; and where we think the Government must increase its legislative ambition to reflect the magnitude and complexity of online risks to children.

The NSPCC's six tests for the Online Safety Bill

The NSPCC has set out six tests that the Online Safety Bill must meet if it is to effectively tackle preventable harm to children. Against each of the six tests, we set out a series of indicators that determine whether regulation is likely to prove effective.

In our updated scorecard, we find that the draft Bill either meets our tests in nine of 27 indicators (or we are broadly satisfied with the Government's proposed approach.) However, in a further ten indicators, our tests remain largely or wholly unmet.

Our unmet indicators reflect particular concerns in a number of areas: the absence of an overarching general safety duty; the illegal safety duties inadequately addressing the cross-platform nature of risks; an insufficiently proactive response to content that directly facilitates illegal content; the lack of senior management liability for substantive compliance with the illegal and child safety duties; and the need to introduce statutory user advocacy for children, funded by the industry levy.

---

[1] Figures from an NSPCC Freedom of Information request of police forces in England and Wales. Police Scotland figures show indecent communication with a child offences increased by 80 per cent in the most recent five years.

Safety duties and a systemic approach to harm

The Online Safety Bill must deliver a well-designed, proportionate regulatory framework that results in the strongest possible protections for children. That means the adoption of a systemic, principles-based approach to regulation, underpinned by a broad future proofed Duty of Care.

Although we are satisfied the draft Bill proposes a largely systemic approach, we are surprised that it does not propose an **overarching general safety duty**. An overarching duty could effectively 'sit above' the differential safety duties being proposed. This would provide much-needed coherence to a structurally complex piece of legislation, and it would help ensure the framework of secondary legislation, codes and guidance that it underpins is orientated more clearly around fundamental safety objectives. This could reduce the risk that online services adopt a differential approach to the discharge of their safety duties.

In the model outlined in the NSPCC's regulatory proposal[2], and the original Duty of Care approach set out by Perrin and Woods[3], platforms would be required to identify reasonably foreseeable harms and take reasonable and proportionate measures to address them. The draft Bill proposes that online services should adopt systems and processes to ensure safety that are effective and proportionate (although it is unclear in relation to what.)

This means that platforms would not necessarily be required to take steps that are considered reasonable.

The NSPCC has several concerns about **how the legislation defines harm**, and whether this results in lower standards of protection than comparative regulatory regimes.

For harmful content, material will be considered in scope if there are reasonable grounds to consider there is a material risk of it having, or indirectly having, a significant adverse physical psychological impact on a child of ordinary sensibilities (clause 45(3)). However, this appears to establish a higher threshold for intervention than the recently established rules for Video Sharing Platforms[4], in which children should be protected from material that might impair the physical mental or moral development of persons under the age of 18.

Similarly, the draft Bill introduces a 'child use test' that sets a higher threshold than the ICO's Children's Code in respect of whether a service is likely to be accessed by a child. Clause 26 requires that a 'child user condition' is met, and that a service is only considered as being 'likely to be accessed by children' if there are a significant number of children using it, or the service is likely to attract a significant number of child users.

The definition of 'significant' is not adequately set out, but this raises the prospect that many smaller or specialist sites could be excluded from this part of legislation. Problematic platforms including Telegram and OnlyFans might legitimately argue either that their predominant user base is adults, or

---

[2] NSPCC (2019) Taming the Wild West Web: How to regulate social networks and keep children safe from abuse. London: NSPCC

[3] Perrin, W and Woods, L. (2019) Internet Harm Reduction: a proposal. Dunfermline: Carnegie UK Trust

[4] Ofcom (2021) Consultation: guidance for video sharing platform providers on measures to protect users from harmful material. London: Ofcom

that even a substantive minority of child users nevertheless falls below the qualifying threshold set out in the regime.

As a result, there is a considerable risk that harmful content may simply be displaced to smaller sites and those outside of scope.

There is a potential systemic weakness in the risk assessment process, that could in turn ripple throughout the regime. Although the risk assessment process is generally well designed, there appears to be limited means for Ofcom to review risk assessments, nor to take action where the assessments produced are of poor quality.

Given that the scope of obligations to tackle harmful content largely stems from how and what risks are perceived in the first place, the legislation introduces a risk of moral hazard for online services to overlook the more risk-inducing or complex aspects of their services.

It remains unclear whether an assessment of harm is to be made by considering the impact of an individual piece of content, or the cumulative impact of such pieces of content (including the impact of this being algorithmically recommended to children.)

The effectiveness of the child abuse response

We strongly welcome the clear emphasis in the draft Bill on tackling technology facilitated sexual abuse, with all online services subject to a safety duty in respect of illegal content.

Platforms will be required to risk assess their services; use proportionate systems and processes to effectively manage the risk of harm to individuals; and to minimise the presence of priority illegal content, the length of time for which is present, and how easily it will be disseminated.

The regulation should clearly and unambiguously require online services to adopt a consistent and effective child abuse response. This must include, but not be limited to, the scope and effectiveness of their takedown processes; measures to proactively detect and disrupt new images being produced; and mechanisms to proactively detect and report online grooming.

Much will rest upon the scope and ambition of Ofcom's approach, and whether it is willing to adopt a regulatory scheme that is commensurate to the scale and extent of online abuse. Ofcom should develop its regulatory scheme with a clear understanding that a satisfactory response will likely need to exceed the action currently undertaken by many sites. The regulator should avoid a default assumption that the current approaches of the larger firms are the upper limits of what is required.

In a number of substantive areas, we have concerns about whether the legislative framework adequately responds to the scale and dynamics of the child abuse threat.

Firstly, the draft Bill fails to adequately respond to the **cross-platform nature of online abuse**. Online sexual abuse is rarely sign on a single platform or app: for example, there are well-established online grooming pathways, in which abusers exploit the design features of social networks to make effortless contact with children, before the process of coercion and control over them is migrated to encrypted messaging or live streaming sites.

If the regulatory regime is to be effective, it will require a systematic response to cross-platform risks. No one online service can assemble every piece of the jigsaw. Platforms have already

demonstrated this is achievable, albeit primarily through targeted and largely content focused initiatives.[5]

At present, the draft legislation is at best unclear about the requirements to consider cross-platform risks. For example, the risk assessment process for illegal content refers to material encountered 'by means' of the service, but doesn't specify whether this relates only to content encountered on the site, or the ways in which activity on other platforms could contribute to illegal material being accessed. It is also poorly structured to address the constraining influence of competition law on any potential harm reduction approaches.[6]

We strongly encourage the Government to amend the legislation to introduce a duty on Ofcom to address cross-platform risks. Online services should have a clear duty to co-operate on the cross-platform nature of risks; risk assess on that basis; and demonstrate compliance with their safety duties through an active assessment and mitigation approach for cross-platform risks.

The draft Bill inadequately addresses the growing challenge of **content that facilitates illegal behaviour, but that may not in and of itself meet the criminal threshold** for removal. Unless the Online Safety Bill gives the regulator powers to treat content that facilitates child abuse with the same severity as illegal material, through amending the scope of the illegal safety duty, legislation will fail to tackle egregious material upstream. A crucial opportunity to prevent and disrupt abuse at an early stage will be lost.

Abusers will still be able to organise in plain sight; post 'digital breadcrumbs' that signpost to illegal content; and continue to re-victimise children through the sharing and viewing of carefully edited child abuse sequences[7].

Although we strongly endorse the Government's decision to include both public and private messaging in the scope of the Bill, we are concerned about the potential effectiveness of the proposed approach to address **child abuse risks in private messaging**.

Recent data from the Office for National Statistics (ONS)[8] shows that private messaging plays a central role in contact between children and people they have not met offline before. When children are contacted by someone they don't know in person, in nearly three quarters (74 per cent) of cases, this contact initially takes place by private message.

Under the draft Bill, the regulator be able to address would such risks by being able to compel platforms to use automated technologies to detect child abuse content, on both public or private parts of its service, through the use of a 'technology warning notice' (clause 63.)

---

[5] For example, there are cross industry approaches to identify and remove child abuse and terrorist content. TikTok has proposed an industrywide scheme to identify and takedown harmful content, aimed at preventing the speed with which content can proliferate

[6] Any activity to tackle cross-platform risks will need to be structured to ensure there is no negative interplay with the Competition Act 1998. This is best achieved through an explicit duty in the Online Safety Bill. Alternatively, the Secretary of State could make an order to exclude cross-platform co-operation from Chapter 1 of the Competition Act, although such orders are normally used for time-limited matters, such as to support supply chains and transport provision during the pandemic.

[7] In many cases, offenders use such edited sequences to 'game' content moderation rules, being able to build up a detailed understanding of what content will not be proactively removed by the host site. Canadian Centre for Child Protection (2019) How We Are Failing Children: changing the paradigm. Winnipeg: C3P.

[8] Office for National Statistics (2021) Children's online behaviour in England and Wales: year ending 2020. Newport: ONS

While we support the principle of technology warning notices being used in a proportionate way, we are concerned that the proposed process sets a very high bar before regulatory action could occur. In practice, it might be highly challenging for the regulator to exercise these powers. This is because:

- the regulator will need to demonstrate the prevalence and persistent presence of child abuse content before it can issue a technology warning notice. This seems to run contrary to the proactive and upstream emphasis on harm reduction set out elsewhere in the legislation;
- the proposed approach presents a potentially unresolvable Catch-22: there are significant questions about how and whether such a threshold can be met, when design choices such as end-to-end encryption would result in a steep fall in reporting volumes and capability;
- Ofcom would need to be satisfied that a platform has failed to address persistent and prevalent abuse, but companies might be able to offset this risk by reporting superficially high metrics that may be suggestive of a highly effective response, but that cannot easily or readily be understood in the context of the actual magnitude of abuse taking place.

We are concerned that this aspect of the legislation is not future proof. Sites including Twitter are actively developing proposals to move to a decentralised operating model, which would effectively 'engineer away' the ability to perform content moderation altogether (and in turn comply with this part of the legislation.)[9] Under such circumstances, Ofcom would have relatively little leverage to secure compliance.

We encourage the committee to explore with Ofcom its assessment of which automated technologies that are currently in use, it envisages could form part of an approved list to be used in technology warning notices. As a minimum, Ofcom should envisage hash scanning, and visual and text based classifiers, as part of its approved set of technologies.

Achieving a higher standard of protection for children

We welcome the draft Bill's ambition to provide a higher standard of protection to children than adults. The legislation must tackle clearly inappropriate and potentially harmful content. This includes material that promotes or glorifies suicide and self-harm, which most major sites prohibit but often fail to moderate effectively.[10] In many cases, the potential harm is likely to come from mechanisms that promote or algorithmically recommend harmful content to users.

The most serious legal harms continue to affect children at scale, and in response to rapid technological and market changes, new harms may quickly emerge and the impact of substantive risks to children may increase. Although some have argued that harmful content should be addressed wholly through changes to the legal framework rather than through regulatory ends[11], we

[9] Twitter has formed a new unit, Blue Sky, which aims to develop a decentralised standard for social networks, staffed with crypto and blockchain expertise

[10] For example, Facebook's transparency reports suggest that up to five in 10,000 views contain prohibited suicide and self-harm content. For vulnerable users being algorithmically recommended content, this is likely to be much higher

[11] For example, the Lords Communications and Digital Committee recently measures to this effect. Lords Communications and Digital Committee (2021) Freedom for all? Freedom of expression in the digital age. 1st Report of session 2021-22

cannot conceive how such an approach could provide children with appropriately future proofed or adequately agile protections.

It seems likely that the legislative provisions to protect children will be contingent on platforms rolling out **age assurance technologies**. Such technologies would determine with reasonable certainty whether a user is a child, and therefore requires additional regulatory protections as set out in the regime.

Given the intrinsic role of age assurance to deliver a higher standard of protection for children, the Government should set out further detail about how it envisages age assurance being implemented. Further clarification is required about if and when it intends to set standards for age assurance technologies. While the ICO intends to publish further guidance on age assurance measures later this year, it remains highly unclear what standards and thresholds are likely to apply.

If age assurance technologies cannot be rolled out as intended, particularly among smaller platforms that might find harder to develop solutions to the necessary standard, the Government should set out how it envisages that its policy and regulatory objectives can be met.

In parallel to the development of the Online Safety Bill, the **Law Commission has proposed a number of substantive changes to the legal framework**, including a new harm-based communications offence, intimate image-based offences[12], and an offence of encouraging or assisting serious self-harm.[13]

Although these proposals are welcome, not least because criminal law has not kept pace with growing risks of technology facilitated abuse, these changes are likely to have significant implications for Ofcom's regulatory regime. Substantial areas of harm, including material that facilitates child sexual abuse and that encourages or incites self-harm, could be reclassified as relevant criminal offences and therefore subject to the illegal safety duties rather than child or adult safety ones.

This is a product of the unnecessary structural complexity of the Bill, and should at least in part be addressed through adopting our recommendations for an overarching safety duty, and an expansion of the illegal safety duty to cover activity that directly contributes to, or results in, priority illegal content.

In any event, the Government should address how it intends to manage the significant levels of ambiguity, and the resulting challenges this poses for effective parliamentary scrutiny and the development of and compliance with the regulatory regime.

We welcome the Secretary of State's comments to the Committee that he is open to resolving the **exclusion of many commercial pornography sites from regulatory scope**.[14] As it stands, pornography sites would be fall outside of regulatory requirements if they do not host user generated content, or are repurposed to that effect.

Unless the legislation is amended, the Online Safety Bill would therefore offer less protection against age inappropriate content than either the Digital Economy Act or the UK Video Sharing Platform regulations. (Ofcom's VSP regime requires explicit age verification measures for services that host

[12] Law Commission (2021) Intimate image abuse: a consultation paper. London: Law Commission
[13] Law Commission (2021) Modernising communications offences: a final report. London: Law Commission
[14] Comments made by the Secretary of State Oliver Dowden in an oral evidence session to the Digital, Culture, Media and Sport Select Committee, held May 13th 2021

pornographic or sexually explicit content, but only apply to a relatively small number of UK-based services.)

Ofcom's investigatory and enforcement powers

Effective investigatory, information disclosure and enforcement powers are crucial to the regulator's work. The draft Bill should provide Ofcom with a comprehensive set of powers to ensure it can effectively understand how and whether platforms are complying with their safety duties. It is also essential that the regime appropriately incentivises platforms to comply with the legislation, and that the regime seeks to embed a culture of compliance within regulated firms.

We are pleased that Ofcom will be able to launch investigations with a broad range of powers at its disposal, with the ability to issue information requests, powers to interview staff, and commission a Skilled Persons report. We also welcome Ofcom's information disclosure powers being applied to relevant ancillary bodies, which could include app stores or third-parties that support platforms to discharge their regulatory responsibilities.

The draft Bill makes provision for a duty on platforms to publish annual transparency reports, for each of the three safety duties which are applicable. Clause 49 sets out broad parameters for transparency reports. In practice, the scope and effectiveness of transparency reporting will largely be determined by how Ofcom decides to implement this part of the regime. The Bill's impact assessment sets out 10-year compliance costs for transparency of only £3.6 million[15], which raises concerns whether a relatively limited set of transparency measures may be sought.

We are disappointed that the draft Bill fails to introduce broad but workable **information disclosure duties** on platforms. Experience from the financial services sector demonstrates the importance of disclosure duties as a means of actively hardwiring regulatory compliance into senior manager and corporate decision-making.

We encourage the Government to revisit its approach to information disclosure duties. In particular, we recommend that category one services should be covered by a proactive duty to disclose to the regulator information about which it could reasonably expect to be made aware.

Although potentially broad, financial services regulation demonstrates that the scope of this duty can be drawn with sufficient clarity that platforms can properly understand their requirements, and do not face unmanageable reporting burdens.[16]

We have significant concerns about the Government's proposed enforcement approach. The decision not to introduce **broad-based senior management liability** is a significant missed opportunity to incentivise regulatory compliance, and to actively embed the discharge of illegal and child safety duties into corporate decision-making.

The draft Bill makes provision for reserved powers to introduce criminal sanctions against senior managers, but these proposals seem poorly targeted towards delivering child safety outcomes:

---

[15] UK Government (2021) Online Safety Bill - Impact Assessment. London: UK Government

[16] Companies should also be subject to 'red flag' reporting requirements, in which they would be required to disclose to the regulator any significant lapses in their systems and processes that could affect their discharge of the safety duties

sanctions would only apply in circumstances where a senior manager fails to comply with an information request, or knowingly seeks to mislead. Crucially, they would not apply in respect of actual product or safety decisions.

As a result, there is no direct relationship in the Bill between senior management liability and the discharge by a platform of its safety duties.

Based on the experience of other regulated sectors - principally financial services - there is a compelling case for both corporate and senior management liability. The Bill should introduce a Senior Managers Scheme that imposes personal liability on staff whose actions consistently and significantly put children at risk.

Senior managers exercising a 'significant influence function' should be subject to a set of conduct rules that incentivise senior managers to internalise their regulatory requirements when setting business strategy and taking operational decisions. Under such a scheme, the regulator could bring proceedings against senior managers that breach their child safety duties, with proportionate sanctions such as fines, disbarment or censure.

For the most significant failings, there should be provision for criminal sanctions, but only where there is a clear evidence of repeated and systemic failings that result in a significant risk of exposure to illegal harm. Such an approach is wholly consistent with existing jurisprudence relating to systemic failures of duties of care.

User advocacy arrangements

Effective user advocacy arrangements will be integral to the success of the regulatory regime. The draft Bill doesn't include user advocacy provisions, but we welcome the Government's commitment to bring forward proposals during pre-legislative scrutiny.

The legislation should deliver bold and ambitious user advocacy proposals, including a **statutory user advocacy body for children, funded by the industry levy**. This is essential to create a level playing field for children - to ensure there is an effective counterbalance to industry interventions, and to provide regulator with credible and authoritative evidence, support and challenge.

At present, a range of civil society organisations represent children. However, it should not be taken for granted that civil society and charitable organisations can continue to perform these activities in perpetuity, or to the level and extent that is required.

Tech firms are a well-resourced and powerful voice, and will legitimately seek to exert strong influence when decisions are made about their services. Powerful industry interests are not unique to the tech sector, but the size of and resources available to the largest companies are arguably distinct.

In most other regulated markets, these risks are addressed through strong, independent advocacy models[17]. Without such arrangements in place for online harms, there is a clear risk that children's

[17] For example, Citizens Advice acts as the statutory watchdog for the energy and postal markets, the Consumer Council for Water represents water users, and Passenger Focus represents rail and bus users. The value of funded user advocacy arrangements is set out well by Citizens Advice in their assessment of sectoral arrangements. Citizens Advice (2018) Access Denied: the case for strong protections for telecoms users.

interests will be asymmetrical to those of industry, and unable to compete with their worldview and resources.

Creating a level playing field for children means drawing more directly on what exists in other regulated settlements, from postal services to public transport, where the user voice is funded and empowered. Children are potentially the most vulnerable of all users, and they deserve the strongest possible set protections.

Children, at heightened risk of sexual abuse online, should receive no less statutory user advocacy protections than users of a post office or passengers on a bus. To that end, the industry levy is an appropriate mechanism for funding such user advocacy arrangements - it is entirely consistent with the well-established 'polluter pays' principle.

Ensuring Ofcom's effectiveness

The draft Bill is essentially a framework, in which a range of secondary legislation, guidance and codes of practice will sit.

Given how much of the substantive regime will be developed by Ofcom, rather being set out in the primary legislation, it is therefore important that the regulator has the resources and expertise available to act credibly and effectively; that its independence is safeguarded throughout; and that there is extensive and ongoing Parliamentary scrutiny of its plans and performance.

We therefore encourage the Committee to closely scrutinise the proposed interplay between the Secretary of State and regulator, in particular the powers open to the Secretary of State to issue a statement of strategic priorities (clause 109) and to amend a Code of Practice to 'reflect Government policy (clause 33(1).)

The Committee should ensure that the regulator is demonstrating the necessary ambition and expertise to meet its statutory obligations to ensure children receive a higher standard of protection than adults, and that the regulator is willing and able to act decisively to protect children from inherently preventable harms.

We particularly encourage the Committee to explore Ofcom's developing understanding of the online child abuse threat, and to be satisfied that it is demonstrating a clear, effective and consistently child-centred approach to the discharge of its functions.