NSPCC

Duty to Protect

An assessment of the draft Online Safety Bill against the NSPCC's six tests for protecting children

September 2021

EVERY CHILDHOOD IS WORTH FIGHTING FOR

Contents

Summary

In May 2021, the UK Government published its long-awaited draft Online Safety Bill. The legislation is an urgent child protection measure – and should be judged on whether it delivers a comprehensive package of measures to prevent inherently avoidable online abuse.

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.

But, as it stands, the legislation needs much greater ambition if it is to go far enough in tackling preventable online harm, and if it is to set a global standard. In the coming weeks, the Joint Committee on the draft Online Safety Bill will have the opportunity to scrutinise the draft legislation, and to ensure its vital legislative objectives are met.

It is essential that the draft Bill tackles the growing scale and complexity of the child abuse threat. Recent NSPCC data shows that online grooming offences in 2020/21 reached a record high – with the number of sexual communication with a child offences in England and Wales increasing by almost 70 per cent in three years.[1] Internet-facilitated abuse has a seen a trend towards more serious sexual offences against children, and the average age of children in abuse material – particularly girls – has trended younger.[2]

The Bill can, and must, protect children from online sexual abuse, and effectively balance the fundamental rights of all users, including children that require a higher standard of systemic protection.[3]

However, if the legislation continues to fall short, children will continue to face entirely avoidable harm. One in five UK internet users will face preventable online abuse, including online grooming and the production and distribution of child abuse images.[4] The unacceptable high cost of industry inaction will continue to be felt by children, families and society.[5]

The NSPCC's six tests for the Online Safety Bill

The NSPCC has led the campaign for a social media regulator – with companies subject to a legally enforceable Duty of Care that requires them to identify reasonably foreseeable risks, and address them through systemic changes to how their services are designed and run.

In conjunction with Herbert Smith Freehills, in spring 2019 the NSPCC published comprehensive proposals for a regulatory model.[6] Last year, we set out six tests that the Online Safety Bill must meet if it is to deliver for children,[7] and to deliver on the Government's ambition to make Britain the safest place in the world to be online.[8]

The NSPCC intends to judge the Online Safety Bill against each of these tests. This report reviews the draft legislation and sets out our current assessment of whether the tests are being met. In our scorecard (appendix one), we find that while the Government's response sets out a broadly workable and robust regulatory model, there are a number of significant weaknesses which need to be addressed.

Against each of the six tests, we set out a series of indicators that will determine whether regulation goes far enough to protect children from avoidable abuse.

In nine out of 27 indicators, we find the Government has met our tests (or we are broadly satisfied with the proposed approach). However, against a further ten indicators, our tests have been largely or wholly unmet.

1 NSPCC data from a freedom of information request to police forces in England and Wales, August 2021
2 Salter, M; Whitten, T (2021) A contemporary analysis of pre-Internet and contemporary child sexual abuse material. Deviant Behaviour, forthcoming
3 For a more detailed discussion of how online services should balance user privacy and safety considerations, see NSPCC (2021) Private messaging and the rollout of end-to-end encryption: the implications for child protection. London: NSPCC
4 Data from the Information Commissioners Office
5 The Center for Humane Technology maintains a ledger harms that lists the 'negative impact of social media that do not show on the balance sheets of companies, but of society'.
6 NSPCC (2019) Taming the Wild West Web: How to regulate social networks and keep children safe from abuse
7 NSPCC (2020) How to Win the Wild West Web: Six tests for delivering the Online Harms Bill. London: NSPCC
8 UK Government (2019) Online Harms White Paper

What needs to change?

If the Online Safety Bill is to fully deliver for children, the Government should adopt a more ambitious and child-centred approach in some crucial areas of the Bill. In particular, the Government should:

Introduce an overarching general safety duty: An overarching safety duty could effectively 'sit above' the differential safety duties being proposed, and provide much needed coherence to a structurally complex piece of legislation. Crucially, it would help ensure the framework of secondary legislation, codes and guidance that will come together to form the online safety regime are tightly focussed around the Bill's fundamental safety objectives. This could also reduce the risk that online services adopt a differential approach to the discharge of their safety duties.

Ensure regulation addresses the cross-platform nature of risks: Well-established grooming pathways see abusers exploit the design features of social networks to contact children, before they move communication across to encrypted messaging and livestreaming sites.[9] Similarly, harmful content spreads with considerable velocity and virality across social networks and messaging sites.

Unless the Online Safety Bill more effectively responds to the dynamics of the abuse threat, its overall effectiveness will inevitably be constrained. Ofcom must therefore have a legal duty to address the cross-platform nature of risks; with clear expectations on companies that meeting their illegal content and child safety duties means having processes in place to assess and respond to the risks of cross-platform harms, and sharing data on offending behaviour and highly agile and constantly evolving threats.

Take a clearer and more robust approach to activity that directly facilitates online child abuse which may not meet the criminal threshold. 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 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.[10]

Adopt a strengthened approach to tackling harmful content for children: The Bill intends to offer a higher standard protection to children than adults, but introduces a 'child use test' which sets a higher threshold than the ICO's Children's Code in respect of whether a service is considered likely to be accessed by a child. As it stands, this means highly problematic services including Telegram and OnlyFans could potentially be excluded from this part of the legislation, because they could legitimately claim that children don't account for a 'significant' part of their user base. This is likely to result in lower standards of overall protection, and the risk that harmful content is simply displaced to sites not covered by the child safety duty.

Develop a more effective, future-proof response to private messaging risks: Although we welcome the Bill's scope including both public and private messaging, we have concerns about whether its proposed approach, the use of technology warning notices, will prove effective. The regulator will be able to require companies to use automated technology to scan for child abuse images, through issuing a technology warning notice, but can only do so where there is demonstrable evidence of persistent and prevalent child abuse. However, it remains unclear how and whether this evidentiary threshold could ever be met – particularly if design choices such as end-to-end encryption and decentralised operating models eliminate or substantially weaken reporting volumes and detection capability.

Strengthen the proposed enforcement regime through introducing senior management liability that is directly linked to the discharge of the Duty of Care. Under the draft Bill, senior managers will no longer be held personally accountable for decisions on product safety, only for narrow procedural failings. Even then, the powers would not be enacted until at least two years after regulation takes effect.

9 Europol (2020) Internet organised crime threat assessment. Lyon: Europol

10 Drawing on their often-sophisticated knowledge of platform content moderation arrangements to perpetuate access to illegal abuse images and 'game' how content rules are interpreted. Canadian Centre for Child Protection (2019) How we are failing children: changing the paradigm. Winnipeg: C3P

There remains an overwhelming case for a Named Persons Scheme, similar to the highly effective arrangements in financial services, with senior managers subject to fines, disbarment and censure for systemic failures to protect children. For the most significant failings, there should be criminal sanctions, but only where there is clear evidence of repeated and systemic failings that result in a significant risk of exposure to illegal harm.

Deliver a bolder and more ambitious approach to user advocacy: The Government must commit to a dedicated user advocacy voice 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 provide the regulator with credible and authoritative evidence, support and challenge.

The draft Bill should draw more directly on what exists in other regulated sectors, from postal services to transport, where the user voice is funded and empowered. As it stands, children – the most vulnerable of internet users, and at clear and heightened risk of online sexual abuse - will receive less systemic advocacy protections than customers using a post office or passengers on a bus.

Our six tests for the Online Safety Bill

1. Regulation must have, at its heart, an expansive **principles-based duty of care**, capable of driving cultural change;
2. Regulation must meaningfully **tackle child sexual abuse;**
3. The Duty of Care must meaningfully address **legal but harmful content,** both content and how it is recommended to users;
4. There should be **effective transparency requirements and investigation powers** for the regulator, with information disclosure duties on regulated firms;
5. We need to see an **enforcement regime capable of incentivising cultural change**, which should include senior management liability, and criminal and financial sanctions; and
6. There needs to be **user advocacy** arrangements for children, including a dedicated user advocate voice, funded by the industry levy, so children have a powerful voice in regulatory debates

Test one: The Duty of Care

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

Against this test, we are broadly satisfied that the Government envisages a systemic approach to regulation. In our scorecard, measures have been partially or fully met (or, in a number of cases, will be determined by how Ofcom develops its regulatory scheme).

However, we strongly encourage the Government to reconsider its omission of a general safety duty. This overarching duty would give coherence to an otherwise highly complex regulatory regime, and ensure the framework is tightly focussed around its fundamental safety objectives.

Systemic approach to safety duties

In the model outlined in the NSPCC's regulatory proposal,[11] and the original Duty of Care approach set out by Perrin and Woods,[12] platforms would be required to identify and act on activity which present a reasonably foreseeable risk of significant adverse physical or psychological harm to children.

Companies would be required to understand the risks to individuals using their services, including those that result from their design and operation, and put in place appropriate systems and processes to improve safety and monitor their effectiveness.

Although the draft Bill proposes a largely systemic approach, it does not propose an overarching general safety duty. Instead, there are three thematic duties of care, with duties applying in relation to illegal content (clause 7 and 9); if likely to be accessed by children (clause 10); and if large or high-risk services are likely to be accessed by adults (clause 11).

For each duty, relevant platforms will have to identify risks and take proportionate steps to mitigate them ('safety duties'). Each differential duty is accompanied by underpinning obligations to perform a risk assessment.

We strongly encourage the Government to revert back to a general safety duty, as it initially proposed in its Online Harms White Paper. As recommended by Woods et al in their excellent analysis of the draft Bill,[13] this 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 to ensure the framework focuses clearly around its fundamental safety objectives. In doing so, this could reduce the risk that online services adopt a differential approach to the discharge of their safety duties.

Regulatory scope and the definition of harm

The regulatory regime will be broad in scope, encompassing 'user to user' services and search engines. 'User to user' services are defined as sites that host user generated content, and will be in scope if:

- they have a significant number of UK users;
- the UK is a target market for the service; or,
- there are reasonable grounds to believe the service presents a material risk of significant harm to UK individuals.

All online services in scope will be required to tackle illegal content. If platforms are likely to be accessed by children (and have a significant number of child users), they will also be required to act to prevent exposure to harmful content.

Content must meet certain thresholds to be considered harmful, and therefore in scope of the regulation. For illegal content, relevant offences will include child abuse and exploitation; offences where the intended victim is an individual; or where within certain parameters, the Secretary of State specifies the offence falls within scope.

Under clause 41(5) the Secretary of State will have powers to designate 'priority illegal' forms of content, which should be considered a priority for the regulatory regime. Child abuse material is not automatically considered priority content, but we would expect it to be designated as such. Content will be considered in scope where the service provider has 'reasonable grounds' to believe content is illegal.

For harmful content, it will be considered in scope if there are reasonable grounds to consider there is 'a material risk of the content having, or indirectly having, a significant adverse physical or psychological impact on a child of ordinary sensibilities' (clause 45(3)). This appears to establish a higher threshold for intervention than the recently established framework for Video Sharing

11 NSPCC (2019) Taming the Wild West Web: How to regulate social networks and keep children safe from abuse. London: NSPCC
12 Perrin, W and Woods, L. (2019) Internet harm reduction: a proposal. Dunfermline: Carnegie UK Trust
13 Woods, L; Perrin, W.; Walsh, M (2021) The Draft Online Safety Bill: Carnegie UK Trust initial analysis. Dunfermline: Carnegie UK Trust

Platform (VSP) regulation, in which children should be protected from material that might 'impair the physical, mental or moral development of persons under the age of 18.'[14]

It also 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 content taken together (including the impact of this being algorithmically recommended to children).

Risk assessments

Risk assessments form an important part of the proposed regulatory framework, and are a crucial part of realising a systemic approach.

Clause 61 requires Ofcom to perform an overarching risk assessment that will underpin much of the regulatory regime, and as part of this, the regulator must prepare risk profiles on different types of online services. This will be a significant undertaking and will inform much of the subsequent development of the regulatory scheme.

It will therefore be important that Ofcom has the resources and expertise available to complete this exercise effectively; that its independence is protected throughout; and that there are appropriate user advocacy mechanisms established by this stage to provide effective counterbalance to industry attempts to influence it, including through direct and indirect means.[15]

Although this is a highly outcome-focused approach, it nevertheless places a significant burden on Ofcom to ensure its risk profile is comprehensive and regularly updated, and that this translates into regularly refreshed codes and guidance.

Once Ofcom has completed this exercise, it must then publish guidance on how platforms should undertake their own risk assessments. A separate risk assessment must be undertaken for each of the relevant safety duties.

For each risk assessment, companies will need to assess the risks of existing services; carry out further risk assessment before making any significant change to their product; and will need to keep risk assessments up-to-date, including when Ofcom makes any significant change to a risk profile.

Under clause 8, new products will need to be subject to a risk assessment prior to launch.

Risk assessments should cover the core characteristics of a service, which includes the user base, business model, governance and other relevant systems and processes. Platforms must also consider the impact of its functionality on the scale and extent of harms, including how its design choices, use of algorithms and the broad operation of its platform may contribute towards the spread of harm.

While the risk assessment approach is generally sound, there appears to be limited means for Ofcom to review risk assessments, nor take action where the risk assessments produced are of poor quality.[16]

Given that the scope of regulatory obligations stems from how and what risks are identified in the first place, it remains unclear how the legislation manages a risk of moral hazard for firms to overlook more problematic aspects of their services (or risk profile).

Effective Codes of Practice

Ofcom will have a duty to issue statutory codes of practice that set out steps companies can take to fulfil their safety duties. This includes a dedicated Code of Practice on online child abuse (clause 29).

Companies may choose to take alternative steps to those set out in the code, as part of an outcome based approach, provided they can demonstrate these are at least as effective. Companies will be judged to be complying with safety duties if they take the steps described in the codes. In relation to illegal content, Ofcom must also be satisfied that child sexual abuse content is not persistently present or prevalent.

The codes of practice are required to be compatible with a set of Online Safety Objectives (clause 30), which are central to the delivery of a safety-by-design approach. The Objectives require regulated services to design and operate their products with systems and processes to ensure safety that are effective and proportionate (although it is unclear how proportionality is to be determined); and that are appropriate to the size of the user base.

Services must provide a higher level of protection for children; and require adequate controls over access to, and the use of, the service by children of different age groups. Platforms must consider the needs of children at different developmental stages.

14 In the UK, Ofcom regulates Video Sharing Platforms, as a result of the Audiovisual Media Services Directive being transposed into domestic law. Ofcom (2021) Guidance for video sharing platform providers on measures to protect users from harmful material. London: Ofcom

15 For a discussion on how tech firms have sought to distort evidence-based understandings of online harms and use third parties to promote their arguments, including academics and NGOs, see for example Abdalla, and Abdalla, M (2021) The Grey Hoodie Project: Big Tobacco, Big Tech and the threat on academic integrity. Preprint. Cambridge, MA: Harvard. Toronto, ON: University of Toronto to our whole

16 Woods, L; Perrin, W.; Walsh, M (2021) The Draft Online Safety Bill: Carnegie UK Trust initial analysis. Dunfermline: Carnegie UK Trust

Test two: tackling online child abuse

The Online Safety Bill will be judged by how effectively it can protect children from online child abuse risks that continue to grow in their scale and complexity. These risks are inherently preventable.

The draft Bill has a clear emphasis on tackling technology facilitated sexual abuse, with all regulated services subjected to a safety duty in respect of illegal content (clause 9). The Government will set out priority categories of offences in secondary legislation, which should include both online grooming and the production and distribution of child abuse images.

Although the draft legislation provides a largely coherent and systemic response, we have significant concerns about whether the draft Bill displays the necessary ambition to respond to the full dynamics of the child abuse threat.

In particular, the legislation needs to more effectively map onto the dimensions of the abuse threat, including the harm ecosystem in which abusers are able to exploit the design features of social networks to groom children and produce first generation child abuse images. Once images have been produced, they can then be shared across messaging platforms, poorly moderated parts of the surface web, and among more sophisticated offenders, on the dark web.

In our scorecard, we find that our test has been only partially met. In four of the eight measures, the Government's proposals fully or at least partially meet our expectations. However, in two key areas, including the importance of adopting a cross-platform approach to risk, and adequately tackling content that facilitates child abuse, our tests remain unmet.

Building an effective child abuse response

The draft Bill introduces an illegal content safety duty, which will require all online services to use proportionate systems and processes to effectively manage the risks of harm to individuals from illegal content; and to minimise the presence of priority illegal material, the length of time for which it is present, and how easily it is disseminated.

Online services will need to complete an illegal content risk assessment, and to comply with one of more Codes of Practice describing recommended steps for complying with the safety duty. Clause 29 requires Ofcom to produce a statutory Code on online CSA, and to ensure this and other Codes are consistent with a set of online safety objectives set out in clause 30.

Legislation must clearly and unambiguously require online services to demonstrate to the regulator the consistency and sufficiency of their child abuse response. This should include, but certainly 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.

The regulator should require platforms to take measures to substantially frustrate the potential for their design features to be readily exploited by abusers. It should also develop its regulatory scheme with a clear understanding that a satisfactory response will likely need to exceed the action currently undertaken by many sites. Ofcom should avoid a default assumption that the current approaches of the larger firms are the upper limits of what is required.[17]

Much will rest upon the scope and ambition of Ofcom's risk profile and codes of practice, the demonstrable exercise of a risk-based approach, and its understanding and willingness to proactively respond to the dynamics of how child sexual abuse materials are produced and shared.

Adopting a cross-platform approach to risk

The draft Bill fails to adequately respond to the cross-platform nature of many online risks to children.

While platforms will be responsible for harms to individuals that happen as a direct consequence of their site, or activity enabled by it, we have significant concerns that the regulatory expectations on firms won't address the ways in which harms typically extend or proliferate across multiple services.

In order to ensure the regulation effectively responds to the dynamics of the child abuse threat, it is essential that the illegal content and child safety duties apply on a cross-platform basis. Online services should have a clear duty to co-operate on the cross-platform nature of child abuse risks, and to risk assess accordingly.

There should be a corresponding duty on Ofcom to assess and act upon the cross-platform nature of harms, including as part of the development of its risk profile (clause 61). Codes of practice should place clear obligations on platforms to share threat assessments, develop mechanisms to share offender intelligence, and ensure a more coherent systemic approach to addressing an online ecosystem in which unmitigated harms could otherwise flourish.

17 For example, some platforms do not appear to adequately enforce child abuse takedown processes. The Canadian Centre for Child Protection, whose Project Arachnid tool has identified 6.1 million child abuse images since 2016, found that some sites routinely refuse to comply with takedown requests of children aged as young as nine or ten. Some platforms argue that if there is any (even very early) signs of sexual maturation, it is not appropriate for them to takedown images, without knowing the age and identity of the child.

Online abuse is rarely siloed 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 livestreaming apps.[18] Harmful behaviour can spread at considerable velocity across social networks and video sharing services.[19] An abuser may be playing video games with a child, while grooming them on an ancillary chat platform, such as Discord.[20]

If the regulatory regime is to be effective, it will require a more systematic response to cross-platform harms. No one online service can assemble all the pieces of the jigsaw. Platforms have already demonstrated this is achievable, albeit primarily through targeted and largely content-focused initiatives, including the deployment of hash databases to identify and takedown child abuse and terrorist content.[21]

Most recently, TikTok has called for an industry-wide scheme to identify and take down harmful content, aimed at preventing the speed at which such content can proliferate.[22]

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 content 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 has been suggested that Ofcom could require platforms to take action to address cross-platform risks, if this is identified as a concern through its risk profile. However, if the extent of cross-platform parameters is not adequately captured in primary legislation, it seems highly unlikely that Ofcom would have the legal or risk appetite to interpret its remit in such a way. More ambitious or comprehensive cross-platform risk mitigations could potentially be challengeable in court.

Interplay with competition law

There is a potential adverse interplay with competition law, with a lack of legislative clarity on how and to what extent platforms can collaborate, potentially acting as a barrier to effective co-operation on cross-platform risks.

Unless these issues are addressed, this could act as a major constraint on the regulatory framework. We identify three main routes to address this:[23]

- The inclusion of a specific duty on platforms to co-operate in the draft Online Safety Bill. This would represent a long-term and durable solution, and would provide Ofcom with a clear basis to develop an effective cross-platform regulatory scheme that is proportionate to the nature and extent of risks;
- The Secretary of State could make an order to exclude cross-platform co-operation from the Chapter 1 prohibitions of the Competition Act 1988. Orders can be made where there are 'exceptional and compelling reasons of public policy',[24] but are usually made to respond to short-term and exceptional circumstances, and as such, are poorly suited to ongoing matters;
- Companies could self-assess that their co-operation on cross-platform issues generates consumer benefits which outweigh any anti-competitive effects. However, in the absence of specific regulatory guidance from the Competition and Markets Authority (CMA) or Ofcom, this provides the least certainty around boundaries of acceptable vs. unacceptable co-operation. This approach may also lead platforms to hide behind this risk to avoid taking robust action, and may lead to inconsistency in approach across services, as platforms appreciably adopt differential risk appetites and compliance strategies.

18 Europol (2020) Internet organised crime threat assessment. The Hague: Europol
19 For example, in September 2020, a graphic video of an act of suicide, first livestreamed on Facebook, spread rapidly on platforms including TikTok and YouTube. Gilbert, D. (2020) Facebook refused to take down a livestreamed suicide, now it's all over TikTok. New York City: Vice News
20 Helm, B. (2020) Sex, lies and videogames: inside Roblox's war on porn. New York City: Fast Company
21 For example, the hash lists for terrorism content overseen by the Global Internet Forum to Counter Terrorism (GIFCT)
22 TikTok (2020) TikTok proposes global coalition to protect against harmful content. Blog post on TikTok's website
23 The NSPCC thanks Herbert Smith Freehills for their legal opinion, although the views expressed here are the NSPCC's own.
24 For example, orders were made to support supermarkets to co-operate to overcome supply chain issues and to maintain public transport between the Isle Of Wight and the mainland during the first UK lockdown in spring 2020.

Addressing content that directly facilitates illegal behaviour

The draft legislation fails to adequately address the growing challenge of content that facilitates illegal behaviour, but that may not in and of itself meet the criminal threshold for removal.

Up to now, many online services have been reluctant to shift from a clear but arguably reductionist consensus on the definition and dimensions of the child abuse problem. For the purposes of content moderation, most platforms have adopted an approach where they focus on clearly illegal child abuse material, because it is seen by them to objectively meet a concrete (and therefore easily enforceable) definition.[25]

There is a compelling case this approach does not go far enough, because it fails to adequately respond to the circumstances in which child abuse images are produced, and new and existing images are shared.

Online services should be required to identify and act on images that may not meet the current criminal threshold, but which can facilitate access to illegal images; act as 'digital breadcrumbs' that allow abusers to identify and form networks with each other;[26] and allow children to be actively re-victimised through the sharing and viewing of carefully edited abuse sequences.

In particular, the regulator must be prepared to tackle so-called 'abuse image series'. In many cases, abusers will upload or seek to access material containing large numbers of images taken in the run-up to or following sexual abuse, effectively forming part of a sequence that culminates with images or videos that meet the criminal threshold.

In some cases, these are deliberately used by abusers because they anticipate such images won't be proactively removed by the host site.[27]

Given the clearly egregious nature of such material, and its direct contribution to driving illegal activity, the Government should amend the scope of the illegal content safety duty, granting the regulator powers to treat content that facilitates child abuse with the same severity as illegal material. In turn, this should result in expectations on firms to adopt a more proactive and child centred approach to takedown.

In turn, this would give regulatory certainty to companies that at present either don't do enough, or adopt highly differentiated approaches to this type of content.

This is a proportionate and highly targeted approach, and cannot reasonably be opposed on freedom of expression grounds. It is entirely consistent with the clear, upstream approach advocated by the Duty of Care.

Private messaging and technology warning notices

We strongly welcome the Government's decision to significantly broaden the scope of the draft Bill to include both public and private messaging, and to mitigate the significant adverse impacts of high-risk design features including end-to-end encryption. The Online Safety Bill will not succeed unless its scope includes product features and design choices that pose the greatest risk for children.

Recent data from the Office for National Statistics (ONS) shows that private messaging plays a central role in contacts 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%) of cases, this contact initially takes place by private message.[28]

Some 12 million of the 18.4 million child sexual abuse reports made by Facebook worldwide in 2019 related to content shared on private channels.[29]

End-to-end encryption presents very significant risks to children, because it effectively prevents platforms from being able to identify and disrupt child abuse on their services. In turn, this significantly reduces referrals to law enforcement, and it impedes their ability to investigate offences. For example, the National Center for Missing and Exploited Children (NCMEC) estimates that 70% of Facebook reports could be lost if it proceeds with encryption before appropriate mitigations are in place.

25 According to Evelyn Douek, who notes there is a consensus among industry that the 'desirability and definition of child sexual abuse material is quite properly well settled' and that continual re-evaluation of the child abuse threat is not necessary. However, the definitional parameters are far from settled – for example, the Budapest Convention defines fabricated images as illegal, but the US legal parameters do not, an issue which is likely to become more pressing with technological change. Douek, E. (2020) The rise of content cartels: Using transparency and accountability in industry-wide content removal decisions. New York City: Knight First Amendment Institute, Columbia University

26 Disrupting the formation of abuser networks should be a core objective of the illegal content safety duty. This presents an upstream opportunity to disrupt abuse, and while further research is needed, it seems likely the online formation of abuse networks and forums, has contributed to the trend towards more severe abuse.

27 Canadian Centre for Child Protection (2019) How we are failing children: changing the parading. Winnipeg: CCCP

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

29 Figures from the National Center for Missing and Exploited Children

Under the draft Bill, the regulator would be able to address these 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).

While we support the principle of such powers being deployed in a proportionate way, with appropriate safeguards in place, 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 (clauses 63 and 64). 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 high threshold can be met, when end-to-end encryption is likely to result in a steep fall in reporting volumes;
- Ofcom would need to be satisfied that a platform has failed to address persistent and prevalent abuse. However, companies might be able to offset this risk by reporting superficially high metrics that may be suggestive of a highly effective response, or that cannot easily or readily be understood in the context of the actual magnitude or severity of abuse taking place;[30]
- the regulator will need to be satisfied that no alternative, less intrusive approaches are available. It is unclear what happens if such remedies may be technically possible, for example through on-device hash scanning, but could only be achieved with the cooperation of a third party that is outside of regulatory purview.

We are concerned this aspect of the legislation is not future proof. Sites including Twitter are actively developing proposals to move to a decentralised operating model,[31] which would effectively 'engineer away' the ability to perform content moderation altogether (and in turn comply with this part of the legislation). Under such circumstances, Ofcom would have relatively little leverage to secure compliance.

During legislative scrutiny, it is important that Ofcom provides more information on which automated technologies are currently in use that it envisages could form part of an approved list to be deployed 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.

Approved technologies should target the detection and takedown of known child abuse images; new child abuse imagery (including self-generated material); and the detection and disruption of online grooming.

Given the significant challenges set out above, it remains unclear whether the use of technology warning notices will be able to adequately respond to the risks of new and emerging high-risk design choices.

It would clearly be beneficial for the regulator to be able to take enforcement action at an earlier stage of the regulatory process, where a platform is unable to demonstrate that a high-risk design feature can adequately meet its safety duties. This assessment should be informed by a risk assessment, to be undertaken by the platform, which sets out the likely impact of a high-risk design feature on its future ability to identify and respond to child sexual abuse.

30 Disclosure reporting often tends to emphasise the publication of metrics, but without contextualised information that allows an assessment of the resulting impact and scale of platform response. Douek, E. (2020) The rise of content cartels: transparency and accountability in industry-wide content removal decisions. New York City: Knight First Amendment Institute, Columbia University

31 Twitter has established a new division, Blue Sky, to develop a decentralised social network standard. The unit is headed up by Jay Graber, a cryptography specialist,

Test three: tackling legal but harmful content

The Online Safety Bill 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. In many cases, the potential for harm is likely to come from platform mechanisms that promote or algorithmically recommend harmful content to users.[32]

The most serious legal harms continue to affect children at scale,[33] and in response to rapid technological and market changes, new harms may quickly emerge and the impact of substantive threats may substantively increase. Although some arguments suggest that sufficiently harmful content should be addressed primarily through changes to the legal framework, rather than through regulatory ends,[34] it is difficult to envisage how such an approach could adequately protect users from rapidly changing threats, nor be considered future proof.

The draft Bill aims to offer a higher standard of protection to children than adults. However, there are substantive questions about how effectively the draft Bill can deliver against some important aspects of this important legislative ambition. We are particularly concerned about the 'child use test', which sets a higher threshold than the ICO's Children's Code in respect of whether a service is considered likely to be accessed by a child. This may result in lower protection, and the risk that harmful content is simply displaced to other sites.

Our scorecard reflects these concerns about the Government's proposed approach, with two of our three indicators either fully or partially unmet. Much rests on how Ofcom develops its regulatory scheme, and whether its risk profile and codes of practice adopt a suitably child-centred, harm-based approach.

Achieving a higher standard of protection for children

Under the draft Bill, all online services likely to be accessed by children will have to take proportionate measures to prevent them being exposed to harmful content (clause 9). The Government will set out a list of priority legal but harmful risks in secondary legislation.

Platforms likely to be accessed by children will have to clearly specify what content is deemed acceptable in their terms and conditions. Companies will be required to conduct regular child safety risk assessments, use proportionate systems and processes to prevent children's exposure to harmful content, and have processes in place to monitor their effectiveness.

As part of the risk assessment process, platforms will be required to assess the risk of harms against different age groups. As part of a systemic approach, companies will be expected to take account of the harms identified in Ofcom's risk assessment, and comply with measures set out in codes of practice.

Companies will only have to mitigate risks that have been previously identified in a child risk assessment. As with the other safety duties, the limited means for Ofcom to intervene to set quality measures presents a moral hazard for companies – it may be advantageous to either overlook or only superficially engage with more problematic parts of their services.

Although this risk is somewhat mitigated by the requirements to regularly update risk assessments, and to report on the incidence of new and emerging harms ('non-designated harms'), it is arguably desirable for the safety duties and risk assessment provisions to explicitly require companies to consider all reasonably foreseeable risks.

Differential protections and the 'child use test'

We have significant concerns that the draft legislation introduces a 'child use test', which 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. This may result in lower standards of protection, and given the Bill's failure to appropriately tackle cross-platform risks, is likely to result in significant amounts of harmful content simply being displaced to other sites.

Clause 26 requires that a 'child user condition' is met. Under the clause, a service is only considered as being 'likely to be accessed by children' if there are a significant

32 The ways in which harmful content is algorithmically recommended to young people is set out in 5Rights' recent research. 5Rights Foundation (2021) Pathways: how digital design puts children At risk. London: 5 Rights Foundation. A September 2021 investigation by the Wall Street Journal showed how TikTok's algorithm recommended age inappropriate sexual content to users, including accounts linking to off-site pornographic content on OnlyFans. Stern, J. (2021) How TikTok serves up sex and drug videos to minors. Published in the Wall Street Journal 08/09/21

33 For example, Facebook estimates that up to 5 in 10,000 views may contain prohibited self-harm and suicide content. However, this is likely to be a significant underestimate for vulnerable children being algorithmically recommended similar content. Transparency reports available on Facebook's website

34 For example, the Lords Communications and Digital Committee (2021) Free for all? Freedom of expression in the digital age. 1st report of session, 2021-22

number of children who use 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 possibility that many smaller or specialist sites could be excluded from this part of the legislation, and that the legislation could result in platforms including Telegram and OnlyFans potentially being excluded from regulatory scope.

Platforms could legitimately argue either that their predominant user base is adults, or that even a substantive minority of child users nevertheless falls below the qualifying threshold set.

Taken alongside the qualifying thresholds for harm, which are comparatively higher than those set out in the Video Sharing Platform regulations, in respect of protecting children from any content that might impair the mental, physical or emotional development of under 18s,[35] there is a pronounced risk the draft Bill results in comparatively weaker protections than already provided in existing regulatory regimes.

It remains unclear why the legislation has adopted a higher set of qualifying thresholds than the Children's Code, which clearly sets out that online services are in scope unless they can demonstrate they have effective age verification measures in place.

The ICO intends to adopt a risk-based approach to compliance, which will ensure the regulation is implemented in a highly proportionate way – and that offers comparatively higher standards of protection to children, while also providing additional certainty to companies in regulated scope.

Developing effective age assurance

It seems likely that the legislation will require companies in scope to introduce age assurance technologies, in order to determine with reasonable certainty whether a user is a child, and therefore requires the additional regulatory protections set out in the regime.

Age assurance will be expected to perform much of the heavy lifting to identify children and protect them from age inappropriate or harmful content. During the legislative process, it is therefore essential that the Government sets out further detail about how it envisages age assurance being implemented.

In particular, further clarification is required about whether it intends to set standards for age assurance technologies. While the ICO intends to publish further guidance on age assurance measures later this year,[36] it remains deeply unclear what standards and thresholds are likely to apply.

During the pre-legislative process, it will be important that the Government and Ofcom can demonstrate confidence that its legislative and regulatory objectives can be delivered through age assurance technologies, many of which are still opaque,[37] untested or still to be brought to market.

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

Changes to the wider legal framework

In parallel to the Online Safety Bill, the Law Commission has proposed substantive changes to the legal framework on communications offences. This includes a new harm-based communications offence; an offence of encouraging or assisting serious self-harm;[38] and intimate image-based offences, including an offence of taking or sharing an intimate image without consent.[39]

While these proposals are to be welcomed, not least because criminal law has failed to keep pace with the growing risks of technology-facilitated abuse, the lengthy timescales associated with this work means this will at best be happening simultaneously with parliamentary passage, and alongside the development of Ofcom's regulatory scheme.

Should these offences become law, this is 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, might in future be reclassified as relevant criminal offences - and therefore subject to the illegal content safety duties, rather than child or adult safety ones.

This creates significantly high levels of ambiguity for parliamentary scrutiny, and uncertainty for Ofcom as it tries to develop its regulatory regime. In turn, companies will be unclear of the regulatory requirements that may eventually be expected of them.

35 Ofcom (2021) Guidance for Video Sharing Providers on measures to tackle harmful content. London: Ofcom
36 ICO (2021) As the Children's Code comes in - what next? Blog by Stephen Bonner, ICO Executive Director of Regulatory Futures
37 In August 2021, Instagram announced it would be introducing age assurance measures ahead of the Children's Code taking effect, but provided very limited detail on its proposed approach
38 Law Commission (2021) Modernising Communications Offences: final report. London: Law Commission
39 The final recommendations on intimate image based offences will not be published until spring 2022.

This uncertainty results primarily from the structural complexity of the Bill, and it is through simplification of the proposed regime that this ambiguity should best be managed.

The introduction of an overarching safety duty, accompanied by the illegal content safety duty being expanded to cover activity that directly contributes to or results in illegal harm, would significantly reduce potential ambiguity, and bolster the potential for effective scrutiny of the Bill.

Children's access to pornography

Following the Government's decision not to proceed with part three of the Digital Economy Act, which made provision for age verification for commercial pornography sites, ministers provided reassurance that these objectives would instead be taken forward as part of the Online Safety Bill.

Access to age-inappropriate pornography is a substantive concern: recent research has found that 62% of 11-13-year-olds who reported having seen pornography described their viewing as mostly unintentional.[40]

However, the proposed scope of the legislation continues to exclude many commercial pornography sites. If a pornography site doesn't host user generated content, or is repurposed to that effect, it would no longer be required to protect children from age-inappropriate content.

As it stands, the Bill would therefore offer less protection than either the Digital Economy Act or the UK Video Sharing Platform regulations (the measures will require explicit age verification measures for services that host pornographic and sexually explicit content, but only apply to a small number of UK-based services).[41]

The Government should amend the scope of the Bill to explicitly capture all commercial pornography sites. It could also consider amending the scope of the VSP regulations to encompass commercial pornography sites that target or are accessible by UK users, or explore options as part of its proposed changes to the regulation of audience standards for video on-demand services.[42]

Test four: transparency and investigation powers

Transparency, investigation and information disclosure powers are crucial to the regulator's work. A close relationship between the regulator and regulated firms is essential, which includes transparency and scrutiny on the regulator's terms.

The draft Bill proposes to give Ofcom an effective suite of investigatory powers, although it will be important that the regulator is given the resources it needs to effectively investigate how and whether platforms are complying with their safety duties.

Information disclosure duties could play a valuable role in hardwiring safety duties into corporate activity. It is therefore disappointing the Government has failed to integrate this aspect of regulatory design into the proposed approach, particularly given how effectively this works in financial services,

Our scorecard for this measure is mixed: three out of five indicators are at least partially met, but two – relating to the information disclosure duties - have not been met at all.

Information gathering powers

Ofcom will benefit from comprehensive information gathering powers, with the ability to issue information notices for exercising, or deciding whether to exercise, any of their online safety functions (clause 70). These powers extend to regulated firms and relevant ancillary bodies,[43] which could include app stores or third-parties that support platforms to discharge their regulatory duties.

40 BBFC (2020), BBFC Young People and Pornography. London: BBFC. A detailed understanding of children's access to age-inappropriate material is also set out in Thurman, N. (2021) 'The regulation of internet pornography: What a survey of under-18s tells us about the necessity for and potential efficacy of emerging legislative approaches.' *Policy and Internet*. 1– 18.

41 Ofcom's draft framework for VSP Regulation will require 'appropriate age assurance measures to protect under 18's, including age verification for pornography.' This reflects the requirements in the Audiovisual Media Services Directive that measures to restrict access the content should be proportionate to its potential to cause harm stop

42 DCMS is currently consulting on proposals to extend audience protection standards to video on-demand services.

43 In doing so, mitigating concerns about the so-called 'transparency deficit' in ancillary arrangements such as GIFCT.

The regulator will be able to launch investigations with a range of powers at its disposal. Notably, this includes the ability to commission a Skilled Persons report. Ofcom will have powers to commission a review, and when necessary appoint an independent 'skilled person' to conduct it, with the regulated party being liable for the costs involved (clause 74).

The regulator will also have the power to interview staff (clause 76); and powers of entry and inspection (schedule 5).

Transparency reports

The draft Bill makes provision for a duty on platforms to publish annual transparency reports, for each of the three safety duties which apply.

Transparency reports will prove beneficial if they provide meaningful and interrogable information, compared to existing approaches that have widely been dismissed as a form of 'transparency theatre.'[44] Clause 49 sets broad parameters for transparency reports, including information on the incidence and prevalence of illegal and harmful content; how terms and conditions are upheld; the processes used to deliver safety duties; and how services deliver a higher standard of protection to children.

While the statutory provisions appear generally sound, there are some substantive questions about how these will be enacted. The Bill's risk assessment sets out modest 10-year transparency compliance costs of only £3.5 million.[45] If indicative of Ofcom's likely requirements, this suggests a relatively limited set of transparency measures may actually be sought.

Clause 100 places a duty on Ofcom to publish an annual report that summarises the thematic trends highlighted through industry reporting, and to identify industry best practice.

The legislation allows for some form of quality assurance activity led by the regulator. This will build confidence in the quality and robustness of regulatory disclosures, and minimise the risk that platforms seek to present data in a selective and potentially misleading way.

Proactive information disclosure duties

We are disappointed the Government have not placed broad but workable information disclosure duties on platforms.

Category one services should face regulatory duties to proactively disclose information to the regulator about which it could reasonably expect to be informed about. For example, companies should notify Ofcom about significant changes to their products or services, or to their moderation arrangements, which may impact upon the child abuse threat and its response to it.

A similar proactive duty[46] already applies in the financial services sector. Although potentially broad, the scope of this duty can be drawn with sufficient clarity that social media firms can properly understand their requirements, and that companies do not face unmanageable reporting burdens.

Such companies should also be subject to 'red flag' disclosure requirements, in which they would be required to notify the regulator of any significant lapses in, or changes to, systems and processes that compromise children's safety or could put them at risk.[47] For example, if regulation had been in place over the last 12 months, Facebook might reasonably have been expected to report on the technology issues which it attributes to its sharply reduced detection of child abuse content during the second half of 2020/21.[48]

Experience from the financial services sector demonstrates the importance of disclosure duties to act as an important means of regulatory intelligence gathering; but perhaps more importantly, to provide a useful means of hardwiring regulatory compliance into company decisions on the design and operation of their sites.

44 Douek, E. (2020) The rise of content cartels: transparency and accountability in industry-wide content removal decisions. New York City: Knight First Amendment Institute, Columbia University

45 HM Government (2021) Draft Online Safety Bill Impact Assessment. London: HM Government

46 Principle 11 of the financial services regime

47 Again, similar measures are used effectively in other regulated contexts. For example, financial services companies are required to make reporting disclosures under the anti-money laundering and financial services regime, and licensed gambling firms must report breaches against self-exclusion protocols

48 According to Facebook's transparency reports, two technical problems resulted in the volumes of child abuse content being actioned by the site falling by half during this time period. The reporting only provided limited information about the reasons behind this considerable drop-off, which in turn will result in singficant declines in actionable intelligence being made available to police.

Test five: enforcement powers

If online harms regulation is to succeed, Ofcom must have suitably broad enforcement powers and be able to hold non-compliant sites to account.

This reflects the principle that the platforms that create risks should be responsible for the costs of addressing them. For too long children, families and society have been left to deal with the costs of industry inaction through the devastating emotional, mental and physical (as well as social and economic) costs of child sexual abuse.

However, we have significant concerns that the proposed enforcement approach set out in the final Bill does not go far enough to incentivise compliance. In our scorecard, two of our four indicators remain unmet.

Effective financial penalties

The Bill contains steep financial penalties for firms that breach their regulatory obligations. Clause 85 sets out that companies will now face fines of £18 million or ten per cent of revenue (whichever is higher). Fines of such magnitude will clearly only be levied in respect of the most serious regulatory failings.

Although financial penalties are a crucial part of the proposed enforcement approach, it is questionable whether they offer sufficient deterrent value for the largest tech companies. For companies with significant 'cash in hand', the micro economic effect of fines will be blunted, and are likely to have limited impact best on corporate strategy and senior management behaviour.[49]

Business disruption measures

The legislation proposes a range of ambitious service and access restriction orders, which aim to target non-compliant services through issuing business disruption notices to web hosting services and ISPs, financial services companies and advertising networks (clauses 91 and 93).

These provisions are much more substantive than those set out in previous legislation, for example the ISP blocking powers proposed in the Digital Economy Act. We particularly welcome financial providers being in scope.

In recent months, media coverage about child sexual abuse and human trafficking victims led Visa, MasterCard and Discover to stop processing transactions on PornHub. In turn, this resulted in significant changes to how the platform moderates user generated content, and protects vulnerable users including victims of sexual abuse and exploitation.[50]

Senior management liability

We are strongly disappointed that the draft Bill fails to introduce senior management liability.

This is a significant missed opportunity to incentivise behaviour change in companies that might otherwise continue to put children at risk, and to hardwire the 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: 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.[51] 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.

49 In any event, investigations and appeals can be lengthy, and by time proceedings are concluded business models may have shifted, with fines and legal proceedings simply "priced in" as a cost of doing business. Centre for Data Ethics and Innovation (2020) Online targeting: final report and recommendations. London: HM Government.

50 The Hill (2021) Mastercard updates policy for adult content sellers. Published 14/04/21.

51 Chiu, I (2016) Regulatory duties for directors in the financial services sector, and directors duties in company law – Bifurcation and Interfaces. Journal of Business Law, 2016.

The clear deterrence value, and clear potential for adverse reputational effects, are obvious.

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.

Industry groups have fiercely opposed personal liability. In its final response to the White Paper,[52] the Government set out concerns expressed by tech companies about 'potential negative impact on the attractiveness of the UK tech sector.' As it stands, the Government's proposals are now weaker in this regard than the draft General Online Safety Bill in Ireland, which includes criminal sanctions for both regulatory breaches and a failure to cooperate with investigations.[53] Clearly, tech firms are significantly more important to Irish GDP relative to the UK.[54]

Test six: user advocacy arrangements

Effective user advocacy is integral to the success of the regulatory regime. The draft bill doesn't include user advocacy measures, but the Government has committed to bringing forward proposals during pre-legislative scrutiny.

While this is welcome, the Government needs to be much more ambitious in its plans. On our scorecard, two out of three measures remain unmet.

Creating a strong advocate for children

It is essential the Online Safety Bill makes provision for a statutory user advocacy voice for children, funded by the industry levy. Statutory user advocacy is vital to ensure there is effective counterbalance to well-resourced industry interventions, and to enable civil society to offer credible and authoritative support and challenge.

The regulator is unlikely to deliver the best possible outcomes for children unless there is a strong, authoritative and resourced voice that can speak for children in regulatory debates; can support the regulator to understand often complex child abuse issues; and demonstrate emerging areas of concern at an early stage in the regulatory process.

User advocacy requires the resources and expertise necessary to develop high-quality evidence of a sufficient regulatory threshold. Children are one in five Internet users in the UK – and they need a powerful, consistent and well-resourced voice to cut through on regulatory issues.[55]

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 necessary to support, and where necessary to offer challenge to the regulator.

If there is an inappropriately scaled, poorly focused or insufficiently resourced civil society response, this is likely to significantly weaken the regulator's ability and appetite to deliver meaningful outcomes for children.

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.[56] Without such arrangements in place for online harms, there is a clear risk the children's interests will be asymmetrical to those of industry, and unable to compete effectively with their worldview and resources.

In the development of the online harms regime, there is a delicate balancing act between allowing the proposed regulatory duties to promote innovation, protect free

52 HM Government (2020) Final response to the Online Harms White Paper. London: HM Government

53 the General Legislative Scheme published by the Irish government in December 2020

54 Facebook's own data suggests it added 648 million euros to Irish GDP between 2011 and 2018. Report available on Facebook's website. Digitally intensive sectors were estimated to be worth €44 billion to the Irish economy in 2020; and 2017, accounted for 10.6 per cent of all employment. Technology Ireland (2017) Regs It and the Irish technology sector. Dublin: Technology Ireland.

55 Other regulatory schemes recognise the importance of additional safeguards for children, for example Recital 38 of the General Data Protection Regulation states that 'children merit specific protection as they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to [online services.] Children are also less likely to be able to advocate for their own interests, because of the significant 'cognitive burden' associated with use of empowerment on online services. Centre for Data Ethics and Innovation (2020) Online targeting: final report and recommendations. London: HM Government

56 The value of funded user advocacy arrangements is set out well by Citizens Advice in their assessment of sectoral regulators. Citizens Advice (2018) Access denied: the case for stronger protections for telecoms users. London: Citizens Advice

expression, and to ensure that children - perhaps the most vulnerable of all user groups - are protected.

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 of protections.

Put simply, children who have been or are at risk of sexual abuse should not receive less statutory user advocacy protections than users of a post office or passengers on a bus.

The industry levy is an appropriate mechanism for funding such user advocacy arrangements – it is entirely consistent with the well-established 'polluter pays' principle.

A levy model is a wholly proportionate and reasonable set of costs when considered in terms of the commercial return available to platforms that offer their services to children, but fail to protect them from reasonably foreseeable harms.

Effective supercomplaints process

We support the inclusion of supercomplaint powers in the Bill (clause 106), which will allow 'eligible entities' to raise complaints where the conduct or features of one or more services presents a significant risk of harm to children, or where the complaint is judged to be of particular importance to a large number of users.

Organisations will need to be approved as an 'eligible entity' in order to bring forward super-complaints. More broadly, the super complaints scheme will need to be carefully designed: Ofcom likely intends to set an evidentiary threshold for complaints, which is important to avoid it being flooded with an unmanageable number of submissions, and to deter poor quality and speculative complaints.

The regulator will need to be mindful of the barriers that society organisations face when seeking to build a supercomplaint case, including the financial barriers, and the evidentiary barriers associated with demonstrating harms caused by often opaque design choices, including platform algorithms.

Making financial penalties fund safety outcomes

Clause 87 of the draft Bill sets out that financial penalties must be paid into a Consolidated Fund held by HM Treasury.

However, it would be desirable to build in alternative or complimentary provisions, with the objective that financial penalties should be redirected towards funding online safety initiatives and organisations that directly protect and promote the interests of service users.

An example of existing regulatory arrangements is the Energy Industry Voluntary Redress Scheme, which was established by Ofgem to enable companies subject to enforcement action to make payments to the scheme, in lieu or in addition to a financial penalty that would otherwise be returned to the Exchequer.[57]

Understanding the experience of all children

The regulator should have a specific duty to assess the risk of harm to particular groups of users, and to assess how online harms may be disproportionately experienced by them. This should include a consideration of how online harms may be differentially felt by users with one or more characteristics under the Equality Act.

Provision should be made for the regulator to be informed by a wide plurality of user experience. We recommend that the regulator develop user representation structures for this purpose, enabling it to inform its approach to engagement with those that have experienced online harms, and represent a broad cross-section of UK users (including those that may be exposed to risk on an intersectional basis). These should complement, but not act as a substitute for, high quality funded user advocacy arrangements.

Children are likely to experience online harms in many different ways. For example, there is extensive research which suggests that LGBTQ+ children are likely to face greater levels of harassment and abuse online, and are more likely to be contacted by people online who aren't who they claim to be.[58]

Similarly, the regulator's research programme must ensure it captures the plurality of children's experience.

57 The scheme is operated by the Energy Saving Trust, on behalf of the regulator

58 See for example, McGeeney, E and Hanson, E (2017) Digital Romance: a research project exploring young people's use of technology in their romantic relationships and love lives. London: Brook/CEOP

Appendix one
Scorecard against the NSPCC's six tests

The NSPCC uses a scorecard approach to assess whether the Online Safety Bill and Ofcom's regulatory scheme will meet our six tests for effective regulation. This scorecard sets out the NSPCC's assessment of the draft Bill against these tests.

Against each test, we set out a series of indicators that will determine whether regulation goes far enough to protect children from avoidable abuse.

Key:

- *indicator wholly or largely met*
- *indicator partially met or still to be determined*
- *indicator wholly or largely unmet*

Test one: the Duty of Care

A fully-fledged Duty of Care that requires platforms to take a systemic approach to protecting children, through the identification of reasonably foreseeable harms and proportionate measures to address them

Codes of Practice are intelligently designed, setting out ambitious but deliverable expectations for the discharge of the Duty of Care

Ofcom's regulatory scheme corresponds to the scale of online harms children face, with platforms incentivised to respond to current risks (and notify the regulator of emerging ones)

The Government adopts, as one of the guiding principles for the regulatory framework, an objective for Ofcom to incentivise cultural change through the development of its regulatory scheme

Test two: tackling online child abuse

Ofcom is enabled to deliver a regulatory scheme that requires bold and ambitious action on child sexual abuse

Ofcom demonstrates a clear understanding of the child abuse threat, and emphasises the prevention of avoidable harm is a central focus of the regulatory approach

There are clear and comprehensive expectations on platforms to address how their design features exacerbate child abuse risks, including high risk design features

There are specific requirements to disrupt online grooming, remove illegal content in a child centred and consistent way, and to take steps to prevent the production and distribution of new child abuse images

There is a regulatory duty on Ofcom to address the cross-platform nature of risks, with corresponding requirements on platforms to share data on offending behaviour and threats

The Online Safety Bill ensures an upstream approach to tackling child abuse, with the regulator treating content that facilitates illegal behaviour with the same severity as material that meets the criminal threshold

Private messaging is in scope, recognising it is a major driver for the production and distribution of child abuse images and grooming

The regulator has proportionate but effective mechanisms to address and mitigate the impacts of the highest risk design features, including end-to-end encryption

Test three: tackling legal but harmful content

The regulator develops a comprehensive and highly effective approach to tackling legal but harmful content, recognising its significant impact on children's safety and well-being

Ofcom produces a Code of Practice that clearly sets out what it considers an acceptable response to priority categories of harmful content. This should include moderation strategies, how content is algorithmically recommended to users, and what it considers suitable outcomes from age assurance measures

The scope of the Online Safety Bill is amended to capture all commercial pornography sites

Test four: transparency and investigation powers

The regulator has comprehensive investigatory and information disclosure powers

Annual transparency reports provide meaningful and intelligible information on the scale and extent of abuse risks, and the effectiveness of response

Ofcom is appropriately resourced to conduct thematic reviews and investigations, and has a strong risk appetite for doing so

Category one services face broad but workable information disclosure duties, including a proactive duty to disclose information about which the regulator could reasonably be expected to be aware

Category one services are required to 'red flag' significant breaches of the Duty of Care that compromise children's safety or put them at risk

Test five: enforcement powers

The regulator has a suitable range of enforcement mechanisms for companies, including robust financial sanctions

The regulator is able to use a range of intelligently designed and proportionate business disruption measures

The Government commits to senior management liability that is directly linked to the discharge of the Duty of Care, and that is able to secure the extent of cultural change that is required. Senior managers are personally accountable for decisions on product safety, not only a failure to cooperate with the regulator

Managers exercising a 'significant influence function' are liable for regulatory action if they breach their Duty of Care requirements, with the option of criminal and financial sanctions for the most egregious breaches

Test six: user advocacy arrangements

The Government commits to a user advocacy body for children, funded by the industry levy, to ensure a 'level playing field' for children, and ensure children's interests are represented in regulatory decisions

There is an effective supercomplaints process for systemic breaches of the Duty of Care to be investigated

There should be a duty on Ofcom to assess the risks of harms to particular groups of users, and assess how online harms maybe disproportionately experience by them. This should include an assessment of how online harms may be differentially experienced by users with one or more protected characteristics under the Equality Act.

NSPCC

Everyone who comes into contact with children and young people has a responsibility to keep them safe. At the NSPCC, we help individuals and organisations to do this.

We provide a range of online and face-to-face training courses. We keep you up-to-date with the latest child protection policy, practice and research and help you to understand and respond to your safeguarding challenges. And we share our knowledge of what works to help you deliver services for children and families.

It means together we can help children who've been abused to rebuild their lives. Together we can protect children at risk. And, together, we can find the best ways of preventing child abuse from ever happening.

But it's only with your support, working together, that we can be here to make children safer right across the UK.

nspcc.org.uk

EVERY CHILDHOOD IS WORTH FIGHTING FOR