

NSPCC

Time to act

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

April 2022

Contents

Summary	3
What needs to change?	5
Are our six tests for the Online Safety Bill being met?	6
Test one: The Duty of Care	7
Test two: Tackling online child abuse	10
Test three: tackling legal but harmful content	15
Test four: transparency and investigation powers	18
Test five: enforcement powers	20
Test six: user advocacy arrangements	22
Appendix one: Scorecard against the NSPCC's six tests	24

Summary

The Online Safety Bill has now been published – an urgent child protection measure that should be judged on whether it delivers a comprehensive package of measures to prevent inherently avoidable online abuse.

We strongly support the ambition of the Bill and commend the Government for bringing forward this potentially world-leading legislation. Legislation has the potential to deliver a robust but proportionate regulatory regime, through the adoption of a framework that requires regulated companies to proactively identify and mitigate potential risks to children.

Well-designed legislation will secure the UK's ambition to become the safest place in the world to be online.¹ The Bill can effectively promote safety and free expression objectives, and offer new protections that enable all internet users, including children, to benefit from social networks, gaming platforms and messaging services that are built to be fundamentally safe-by-design.

Since the draft legislation was published, the Government has made welcome changes that strengthen the Bill's protections for children; and the Culture Secretary has repeatedly stressed the importance of improving the legislation to ensure it effectively protects children from online risk.

But further changes are urgently required so the Online Safety Bill can deliver its overarching objective of protecting children from preventable online harm, and to ensure it delivers a fit-for-purpose and upstream approach to tackling online child sexual abuse.

The scale and complexity of online child abuse continues to increase. In this context, we face a unique and vital opportunity to ensure the Online Safety Bill offers a commensurate response to a growing and increasingly severe set of harms:

- 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% in three years;³
- In 2021, UK law enforcement received 97,727 industry reports relating to online child abuse, a 29% increase from the previous year;⁴
- Internet-facilitated abuse increasingly results in more serious sexual offences against children, with the average age of children in child abuse images – particularly girls – trending younger;⁵ 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.⁶

A strengthened Bill will mean that one in five UK Internet users will no longer face preventable online abuse,⁷ including: online grooming; children being coerced into sending self-generated images to abusers; and social networks being used as a conveyor belt to produce new images and to signpost to existing child abuse content.

This legislation must succeed – and the NSPCC will continue to work tirelessly and constructively to ensure that it does. The unacceptably high cost of industry inaction must not continue to be felt by children, families, and society. A stronger Bill will mean children can be protected against inherently preventable online harm.

1 UK Government (2019) Online Harms White Paper

2 Dorries, N (2022) How we will narrow the ground for barring harmful posts in the Online Safety Bill. Published in ConservativeHome, 15 March 2022

3 NSPCC data on a freedom of information request to police forces in England and Wales, August 2021

4 Data provided by the National Centre for Missing and Exploited Children (NCMEC)

5 Salter, M; Whitten, T. (2021) An analysis of pre-internet and contemporary child sexual abuse material. Deviant Behaviour, forthcoming

6 For a more detailed discussion on 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

7 Data from the Information Commissioners Office

The NSPCC's six tests for the Online Safety Bill

The NSPCC has led the campaign for a social media regulator, and through our WildWestWeb campaign, secured Government's initial commitment to introduce the Bill.

Our vision is that tech companies will face a legally enforceable Duty of Care that requires them to identify reasonably foreseeable risks, and to address them through systemic changes to how their products are designed and run.

We have committed to undertaking detailed scrutiny of the Government's legislative plans. Throughout the process so far, we have worked extensively to help shape the proposals and entered into constructive dialogue with ministers, regulators, industry, and civil society to understand how to build a proportionate and highly effective regime.

In spring 2019, in conjunction with Herbert Smith Freehills we published our comprehensive proposals for a regulatory model.⁸ In 2020, we set out six tests that the Online Safety Bill must meet if it is to deliver for children,⁹ and to secure the Government's ambition to make Britain the safest place in the world for a child to go online.

This report sets out our initial assessment of the Online Safety Bill against each of these tests. The report reviews the legislation in detail, and it makes a series of recommendations about how the Bill can be developed to ensure it delivers on the ambition to tackle preventable harm. Where we identify areas where the legislation can more effectively meet its child protection objectives, we outline a series of workable solutions.

In our scorecard (appendix one), we find that while the Government response sets out a broadly workable and robust regulatory model, in a number of crucial areas the legislation would benefit from further improvements. Many of these issues were evident in the draft legislation and require further considered attention during parliamentary passage.¹⁰

Against each of six tests, we set out a series of indicators that will determine whether regulation goes far enough to protect children from avoidable abuse. In four of six tests, the legislation sets out a sound basis for well-designed regulation that stands to improve the safety of children online but could be further strengthened to build in solutions to remaining challenges. In two of six tests more significant changes should be made to ensure the legislation provides a response that meets the scale of the child abuse risk, and we highlight where targeted improvements can usefully be made during parliamentary passage.

⁸ NSPCC (2019) Taming the Wild West Web: how to regulate social networks can keep children safe from abuse. London: NSPCC

⁹ NSPCC (2020) How to Win the Wild West Web: Six Tests for Delivering the Online Safety Bill. London: NSPCC

¹⁰ Many of the NSPCC's outstanding concerns have been shared by a number of Parliamentary committees in their recent scrutiny of the legislation, including the Joint Committee on the Draft Online Safety Bill; the Commons Digital, Culture, Media and Sport Select Committee; and the Commons Petitions Committee.

What needs to change?

If the Online Safety Bill is to fully deliver for children, we suggest a number of areas where the Government should adopt a more ambitious, child-centred and targeted approach.

Our analysis illustrates a number of areas where the response to the risks of child sexual abuse could be made more robust and its efficacy improved. We make a number of developed recommendations to ensure the legislation provides a more effective and fit-for-purpose response to the detection and disruption of a number of threats, including online grooming.

To strengthen the Bill, the Government should:

Introduce duties to tackle cross-platform child abuse: 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 live streaming sites.¹¹ Similarly, harmful content spreads with considerable velocity and virality across social networks and messaging sites.

The Online Safety Bill must effectively respond to the dynamics of the abuse threat to ensure its provisions coherently target the problem. Companies must face clear requirements to tackle the cross-platform nature of harms when meeting their safety duties; risk assess their products to address how they contribute to grooming and abuse pathways; and face a new duty to co-operate on tackling harms across their sites, including through sharing intelligence on rapidly shifting risks.

Tackle the ways in which abuse is facilitated on social networks, but may not meet the criminal threshold: the Bill must effectively tackle the range of ways in which abusers use social networks to form offender networks; post 'digital breadcrumbs' that signpost to illegal content; and share child abuse videos that are carefully edited to evade content moderation guidelines.

This range of techniques, known as 'breadcrumbing', must clearly and unambiguously be brought into scope, to disrupt abusers who currently can organise abuse in plain sight, and exploit social networks to signpost to child abuse content hosted on third party messaging apps, offender forums and the dark web.

By giving the regulator powers to treat activity that facilitates child abuse with the same severity as illegal material, through amending the scope of the illegal

safety duty, legislation will empower Ofcom to tackle egregious harm upstream. Social networks will no longer be able to allow tens of millions of interactions with accounts that actively facilitate the discovery of child abuse material¹² and abuse could be tackled at the earliest possible stage.

Proactively tackle the child abuse risks in private messaging and groups: we welcome the Bill's scope including both public and private messaging. However we have substantive concerns that the legislation places onerous constraints on Ofcom's ability to proactively tackle the significant risks of grooming and child abuse in private messages, and the ways in which abusers share or signpost to child abuse images in private groups.

As it stands, Ofcom would be unable to require any form of proactive technology to tackle child abuse in private messages in its codes of practice, including industry-standard 'hash' tools that are routinely used to detect child abuse images.

Ofcom will need to be equipped to produce a Code of Practice that is capable of responding to the nature and extent of the child abuse threat. If the regulator had to take site-by-site action to address harm after it has already occurred, primarily as a function of regulatory design, the systemic approach to tackling online harms would be weakened.

Adopt a strengthened approach to tackling harmful content for children: the Bill rightly intends to offer a higher standard of protection to children than adults but introduces a 'child use test' that sets a higher threshold than the ICO's Children's Code in respect of whether service is considered likely to be accessed by a child.

This means highly problematic services including Telegram and OnlyFans could be excluded from the child safety duties, because they can legitimately claim that children don't account for a 'significant' part of their user base. This could result in lower overall standards of protection, and harmful content simply being displaced to sites not covered by the child safety duty.

It is unclear which harms to children will be covered by the child safety duties. The Government should therefore commit to publishing schedule of priority harms for children, similar to the list of priority offences in schedules 6 and 7.

¹¹ Europol (2020) Internet organised crime threat assessment. Lyon: Europol

¹² Data suggests there were over 6 million user interactions with certain types of content in Q1 2021, which if annualised suggests there are tens of millions of interactions with such content on surface web sites

Commit to a statutory user advocacy body for children: the Bill should introduce a statutory user advocacy body representing the interests of children, funded by the industry levy. This is essential to create a level playing field: to ensure there is effective counterbalance to industry interventions, provide an early warning function of new and emerging harms, and to provide the regulator with credible and authoritative expertise, support, and challenge.

Legislation should draw more directly on what exists in other regulated sectors, from postal services to public transport, where funded user advocacy models ensure dedicated expertise that can intervene on behalf of users in regulatory decisions. As it stands, children – the most vulnerable of internet users, and clear and heightened risk of online sexual abuse – will receive less systemic advocacy than passengers on a bus or customers of a post office.







User advocacy is a crucial component of building effective regulatory regime and addressing the clear asymmetry with well-resourced regulated companies.

Take steps to hardwire the safety duties, and to deliver a 'culture of compliance' in regulated firms: the Bill would benefit from a number of targeted improvements that would actively promote cultural change in companies and embed compliance with online safety regulations at 'C-suite' and in Board level decision-making.

Senior management liability should be extended to cover substantive product decisions, not simply a failure to cooperate with the regulator. Companies should be required to appoint a senior manager, at or reporting to Board level, who is personally liable for whether a platform meets its safety duties. As it stands, senior managers of wholly negligent companies could escape any personal liability so long as they co-operate with the regulator.

Companies should also face a broader set of compliance responsibilities, including the Joint Committee's recommendation that risk assessments should be approved at Board level. Companies should be subject to proactive information disclosure duties, placing the onus on regulated firms to flag substantive product changes.

Are our six tests for the Online Safety Bill being met?

- 1 Regulation must have, at its heart, an expansive **principles-based Duty of Care**, capable driving cultural change 
- 2 Regulation must meaningfully **tackle child sexual abuse** 
- 3 The Duty of Care must meaningfully address **legal but harmful content**, including how content is recommended and disseminated 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 for product decisions, and financial and criminal sanctions 
- 6 There needs to be **statutory user advocacy arrangements for children**, including a dedicated user advocacy body funded by the industry levy, so children have a powerful voice that counterbalances that of industry 

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 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. Compared to the draft Bill, the approach is more systemic, with significant strengthening of the risk assessment functions, including additional powers for the regulator to ensure the quality and sufficiency of company risk assessments.

However, we remain concerned about the Bill's overall complexity, and continue to assert that a simplified and strengthened Bill would deliver better outcomes for users, including children. Much will also depend on how Ofcom develops its regulatory scheme, including whether it is able to adopt a suitably agile and child-centred approach to increasingly complex harms.

Systemic approach to safety duties

In the model outlined in the NSPCC's initial regulatory proposal,¹³ and the original Duty of Care approach set out by Perrin and Woods,¹⁴ platforms would be required to identify and act on any harms which present a reasonably foreseeable risk of 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 how services are designed and run, and to put in place appropriate systems and processes to improve safety and monitor their effectiveness.

Although the Bill proposes a largely systemic approach, it does not propose an overarching general safety duty. Government also has largely rejected the Joint Committee's recommendations to simplify its structure. Instead, there remains three thematic duties of care, with duties applying in relation to illegal content (clauses 8 and 9); if likely to be accessed by children (clauses 10 and 11); and if large or high-risk services are likely to be accessed by adults (clauses 12 and 13).

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

We remain concerned that the Bill is highly complex, and that this presents unnecessary challenges. For example, contextual child sexual abuse (CSA) is not currently in scope of the Bill and this could weaken its response to tackling illegal harm (as explored in the next section). This stems largely from the decision to proceed with the distinction in the Bill's architecture between illegal and harmful forms of content.

We continue to believe that a simplified Bill would result in better outcomes for children, including through removing ambiguity, and creating a clearer steer for platforms to prioritise the safeguarding of children; supporting civil society groups advocating for users that have experienced harms; contributing towards effective scrutiny during parliamentary passage; and ultimately, through leading to more substantive compliance, enabling companies to more clearly understand their regulatory requirements and how best they should meet them.

Regulatory scope and the definition of harm

The regulatory regime will be broad in scope, encompassing 'user to user' services, search engines, and services that host commercial pornography. '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 prevent children being exposed 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. A provisional list of relevant offences is contained in schedules 6 and 7.

¹³ NSPCC (2019) Taming the Wild West: how to regulate social networks and keep children safe from abuse. London: NSPCC

¹⁴ Perrin, W; Woods, L. (2019) Internet harm reduction: a proposal. Dunfermline: Carnegie UK Trust

Content will be considered harmful to children where it is designated as primary priority content; priority content; or is likely to cause material risk of significant harm to an 'appreciable' number of UK children. This appears to establish a higher threshold for intervention than the recently established Video Sharing Platform (VSP) regime, in which children should be protected from material that might 'impair the physical, mental or moral development of persons under the age of 18'.¹⁵

An effective risk assessment process

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

We are pleased to see the Bill set out a strengthened and more coherent risk assessment process. At the heart of the regime, Ofcom will undertake risk registers and risk profiles for specific types of online services (clause 83). Companies must then use these to undertake their own risk assessment processes. This will be a significant undertaking and will inform much of the subsequent development of the regulatory scheme.

For each risk assessment, companies will need to assess the risks of existing services; carry out a further risk assessment, before making any significant changes to a product or service; and will need to keep risk assessments up-to-date, including when Ofcom makes any significant change to a risk profile. New products will need to be subject to a risk assessment before 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 design choices, use of algorithms and broad operation of its platform may contribute towards the spread of harm.

We welcome the regulator gaining new powers to ensure risk assessments are of sufficient quality, closing a potential moral hazard in the draft Bill that had required companies only to act on risks identified in their risk assessment, and which in turn provided a perverse incentive to overlook more problematic aspects of their services.

However, we recommend that the Bill should require companies to publish or share risk assessments with civil society organisations and proactively with the regulator. Current experience is that companies are unwilling to share risk assessments, even when requested to do so. Transparency will be vital to civil society groups looking to assess and identify any areas where a company may not be meeting its safety duties, and to make full and effective use of the proposed supercomplaints mechanism.

While understandable issues of commercial confidentiality may apply, and there are clear examples where it would not be appropriate nor desirable for any organisation other than the regulator to receive certain types of information, the absence of a requirement on companies to publish their risk assessment seems unnecessarily opaque. This may obstruct scrutiny of regulated companies, and of the broader functioning of the regime itself

We continue to see merit in the Joint Committee's recommendation that company risk assessments should be reported to and signed-off at Board level.¹⁷ This represents a useful means to hardwire the safety duties into the decision-making of regulated companies; and if implemented effectively, could tackle the evident attention deficit in some social networks in respect of children's safety.

We also recommend that the Government commit to introducing an additional statutory Code of Practice covering harms against women and girls, in recognition of the highly gendered nature of online abuse and the disproportionate exposure of girls to online risk.

Securing Ofcom's role and its effectiveness

The Bill is largely a framework piece of legislation, setting out a complex structure in which a range of secondary legislation, codes and guidance will sit.

Much of the regime will ultimately be determined by Ofcom, once primary legislation has been passed, and how it develops its risk profiles and codes of practice. The regime 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.

¹⁵ In the UK, Ofcom regulates Video Sharing Platforms, as a result of the Audiovisual Media Services Directive being transposed into UK law. Ofcom (2021) Guidance for video sharing platform providers on measures to protect users from harmful material. London: Ofcom

¹⁶ For example, in autumn 2021 the NSPCC led a coalition of 60 global child protection organisations in asking Meta to share its impact assessments and data impact assessments relating to child harms, following the revelations made by the whistleblower Frances Haugen. Meta declined on the basis that 'DPIAs [are] living documents which are regularly updated, and therefore are reflective of a data processing activity at a particular point in time'

¹⁷ Joint Committee on the Draft Online Safety Bill (2021) Draft Online Safety Bill: Report of Session 2021-22

Ofcom's ability to understand and proactively respond to highly agile and constantly evolving harms will be key. During the Bill's passage, we encourage Parliament to closely scrutinise Ofcom's approach to tackling child sexual abuse, and the effectiveness of its abilities and mechanisms to capture rapidly evolving harm.

The success of the regime's systemic risk assessment process will be underpinned by, and significantly reliant on, the regulator's ability to rapidly and effectively identify new and emerging harms. In this context, the need for highly effective early warning functions will be key, and the absence of well-defined user advocacy mechanisms in the legislation seems palpable (see test six.)

If Ofcom is unable to rapidly identify new and emerging harms, the resulting delays could mean entire regulatory cycles where harms are not captured in risk profiles or company risk assessments, and an inevitable lag between harms being identified and companies being required to act on them.

During the set-up of the regime, Ofcom will need to establish a significant number of codes – nine are required by statute alone. It therefore becomes appreciably 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 or skew the evidence-base, including through a range of direct and indirect means.¹⁸

Ensuring the regime is future-proofed

Ofcom will be regulating a sector characterised by rapid technological and market change, and it is therefore crucial the regulator has the necessary powers and resources to ensure it can respond to rapidly changing user threats. Analysis undertaken by Carnegie UK suggests that new technologies such as metaverse will be captured by the regulatory regime.¹⁹ We encourage the Government to clarify this is the case, although we are concerned that the reliance upon a set list of illegal harms and lists of harmful content may prove problematic, and unable to capture and respond highly agile and evolving risk profiles quickly.²⁰

Although Ofcom will be able to recommend new categories of content harmful to children, this will be a lengthy process that ultimately requires secondary legislation. In order to respond to the agile and novel type harms likely to emerge on increasingly immersive technologies, there is a compelling case for the regulator to be able to move more swiftly to amend the scope of its regime,²¹ and to ensure regulation works effectively on behalf of service users.

We also remain concerned about the significant powers being made available to the Secretary of State to influence the regulatory regime. These powers present the risk of future interference in ways that could be detrimental to children. For example, it creates an obvious route for regulated companies to lobby future governments to water down more burdensome parts of their requirements.²²

18 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, A; and Abdalla, M (2021) *The Grey Hoodie Project: Big Tobacco, Big Tech and the threat to academic integrity*. Preprint. Cambridge, MA: Harvard; Toronto, ON: University of Toronto

19 Perrin, W; Woods, L (2022) *Regulating the future: the Online Safety Bill and the metaverse*. Dunfermline: Carnegie UK Trust

20 The NSPCC plans further research to understand the potential impact of the metaverse on risks to children, including the ways in which abuse may be perpetrated by victims and experienced by abusers. The experiential nature of the metaverse may present a broad range of implications, including whether existing criminal offences adequately capture the nature of sexual offences on immersive platforms. There is a compelling basis to suggest changes to criminal laws may need to be amended to reflect these new and emergent harms

21 As set out in Carnegie UK Trust's initial analysis of the legislation, published in March 2022

22 We are particularly concerned by the powers available to the Secretary of State to issue a Statement of Strategic Priorities (clause 143) and to amend a Code of Practice for reasons of public policy (clause 40.)

Test two: Tackling online child abuse

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

The Bill has a clear emphasis on tackling online sexual abuse, with all regulated services subject to a safety duty covering illegal content (clause 9). While the legislation sets out a largely coherent and systemic regime, we recommend the Bill is strengthened to ensure it more effectively responds to, and its provisions are commensurate against, the true scale and dynamics of the child abuse threat.

In particular, the Government should strengthen the legislation to unambiguously address the range of ways in which online abuse is actively facilitated on regulated services, and ensure it more effectively maps onto the dimensions of the abuse threat. This includes the ways offenders use social networks to actively facilitate and organise abuse at scale.

We set out ways the Bill should go further to tackle grooming: as it stands, established grooming pathways may not be adequately tackled; the grooming risks in private messaging may not be effectively addressed; and some forms of otherwise preventable harm seem likely to fall out of scope. By strengthening the legislation its overall impact will be significantly sharpened, and the legislation will provide a suitably coherent upstream response to the detection and disruption of online CSA.

Building an effective child abuse response

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

Online services will need to complete an illegal content risk assessment, and to comply with codes of practice covering illegal content and online CSA. Companies will need to set out how they protect users from illegal content in their terms of service; and must specify which if any proactive technology they use to comply with the safety duties.

We welcome the Bill's provision that regulated services must report relevant UK offences to either the National Crime Agency or another body under an alternative reporting regime. In practice, the majority of child abuse reports will continue to be routed through the National Center for Missing and Exploited Children, based in the United States; but this closes off a technical loophole in which UK-based sites such as OnlyFans faced only voluntary reporting requirements.

Effectively addressing cross-platform risk

The Bill must adequately and unambiguously respond to the cross-platform nature of child abuse risks.²³ The Bill should be strengthened to ensure regulation can be effective in tackling both well-established grooming pathways and new and emerging child abuse harms.

Child abuse is rarely siloed on a single platform or app. For example, abusers will look to 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 apps.²⁴ An abuser may be playing video games with a child while actively grooming them on an ancillary chat platform, such as Discord.²⁵

An effective regulatory regime will require a more systematic response to cross platform abuse pathways. 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.²⁶

In order to ensure that regulation effectively responds to the dynamics of the child abuse threat, the illegal content and child safety duties must 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 clear obligations on platforms to share threat assessments, develop proportionate mechanisms to share offender intelligence, and create 'rapid response' arrangements to ensure platforms develop a coherent systematic approach to new and emerging threats.

23 Cross-platform risks have also been highlighted by industry. For example, Meta's response to the Joint Committee highlighted a range of harms that are organised on smaller platforms before being migrated onto its services

24 Europol (2020) Internet organised crime threat assessment. The Hague: Europol.

25 Helm, B (2020) Sex, lies and videogames: inside Roblox's war on porn. New York City: Fast Company

26 For example the hash lists for terrorism content overseen by the Global Internet Forum to Counter Terrorism (GIFCT)

At present, the legislation is 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, and how platform functionality may contribute towards illegal content being disseminated, but doesn't specify whether this relates only to content encountered on the site or elsewhere.

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. Clause 83 enables Ofcom to identify characteristics of services that are relevant to how harms are produced, but because the risk profiles focus solely on the *characteristics* of sites, and not the *dynamics of the risks* themselves, we remain concerned whether these provisions would adequately capture the appreciably cross-platform nature of some types of child sexual abuse, for example grooming pathways.²⁷

If the extent of cross-platform parameters is not adequately captured in primary legislation, this may place a number of constraints upon the regime.

Firstly, it seems highly unlikely that Ofcom would have the legal or risk appetite to interpret its remit to deliver more ambitious or comprehensive cross-platform risk mitigations when these could potentially be challengeable in court.

Secondly, legal advice suggests that cross-platform co-operation is likely to be significantly impeded unless there is a clear statutory basis to enable or require collaboration that might otherwise be impeded by the interplay with competition laws, or because companies to hide behind such risks to avoid taking more robust action.²⁸

Clause 97 provides a clear statutory basis for Ofcom to co-operate and disclose information with overseas regulators for the purpose of regulation and investigations. Logic suggests that a similar statutory basis is needed for platforms to have the confidence and reassurance to create sharing mechanisms, without fear of being in breach of competition law.

Cross-platform risks, new technology and the Digital Markets Act

New and emerging technologies are likely to produce an intensification of cross-platform risks in the years ahead. We are particularly concerned about the child abuse impacts in **immersive VR and AR environments, including the metaverse**. A number of high-risk immersive products are already designed to be platform-agnostic, meaning that in-product communication can take place between users across multiple products and environments.

There is a growing expectation that the metaverse will be rolled out along such lines, with an incentive for companies to design products in this way in the hope it will blunt the ability of governments to pursue user safety or antitrust objectives.

Separately, regulatory measures being developed in the EU, but which are highly likely to impact service users in the UK, could result in significant unintended safety consequences, unless these are addressed by corresponding mitigations in the Online Safety Bill.

The **interoperability provisions in the Digital Markets Act**, while strongly beneficial through a competition lens, will allow communication between users of multiple platforms. Without appropriate safety mitigations in place, this could provide new means for abusers to contact children across multiple platforms; significantly increase the overall profile of cross platform risks; and actively frustrate a broad number of current online safety responses.

For example, companies may no longer be able to use metadata to detect suspicious patterns of grooming behaviour; or they might be able to successfully argue that it is disproportionate to expect them to significantly rework their current threat detection capabilities.

In the absence of any explicit requirement to risk assess and reasonably mitigate cross-platform harms, including harms which may transfer across multiple products either sequentially or simultaneously, these provisions could result in significant unintended consequences, unless necessary mitigations are adopted in the Bill.

²⁷ It has also been suggested Ofcom could use its information gathering powers to inform the development of its risk profiles, but it is unclear what types of information it would be seeking

²⁸ Legal opinion provided to the NSPCC by Herbert Smith Freehills (HSF). NSPCC thanks HSF for their analysis, although the views expressed here are the NSPCC's own.

Material that directly facilitates online sexual abuse

We have significant concerns that the Bill does not address content that directly facilitates online abuse, and it needs to clearly target the range of ways in which offenders use social networks to perpetuate harm.

Unless these techniques, often referred to as 'breadcrumbing' or contextual CSA, are explicitly brought into the Bill, tens of millions of interactions with child abuse material could potentially fall outside of regulatory scope.²⁹

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.³⁰

However, there is a compelling case this approach does not go far enough. Unless regulated companies are made to more effectively tackle the broader ways in which abusers use their services to facilitate abuse, a crucial opportunity to detect and disrupt online abuse will be lost.

At present, abusers use a range of techniques to facilitate abuse on social networks. Abusers produce abuse material 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; and allow children to be actively re-victimised through the sharing and viewing of carefully edited abuse sequences.

In some cases, these are deliberately used by abusers because they anticipate such images won't be proactively removed by the host site.³¹

There is growing evidence about the harm caused by so-called 'tribute sites', in which offenders create online profiles that misappropriate the identities of known survivors. These fraudulent accounts, which typically adopt survivors' names and feature non-harmful

imagery at the account or profile level, are then used by offender communities to connect with like-minded perpetrators, primarily to exchange contact information, form offender networks, and signpost to child abuse material on the dark web.

In Q1 2021, there were 6 million user interactions with content referencing known survivors or commercial websites.³²

There is also growing evidence that abusers are using private groups on Facebook to build offender groups and signpost to child abuse hosted on third party sites. These sites are thinly veiled in their intentions: for example, groups that ostensibly have an interest in children celebrating their 8th, 9th, or 10th birthdays.³³ Recent analysis suggests that Facebook's algorithms recommend are ruthlessly effective at recommending similar sites, through determining the common characteristic of interest (a sexual interest in children.) This even includes recommending similar groups in multiple other languages.³⁴

Recent whistle-blower disclosures have alleged that Meta management are aware of the child abuse problems in Facebook Groups but have failed to develop a coherent response, a situation that is unlikely to change without legislation.³⁵ Meanwhile, Meta continues to decline to act on Facebook Groups where child abuse content is being directly facilitated: despite being notified of the groups, analysis conducted by Professor Lara Putnam suggests that groups comprising over 50,000 members have still not been removed by the platform.³⁶

Other novel ways of signposting to abuse material, including through QR codes, are now also starting to emerge – emphasising the need for a systemic and overarching harm-based approach.

Therefore it is paramount that 'breadcrumbing' is effectively tackled by the Bill. As it stands, companies will not be required to address the risk of their services being used to directly facilitate the discovery of child abuse material, as part of either their illegal content or child safety duties, in part because of how the Bill has been drafted to wholly differentiate between illegal and harmful activity.

29 Annualised data based on interactions with such content in Q1, WeProtect Global Threat Assessment 2021. London: WeProtect

30 According to Evelyn Douek, there is a consensus among industry that the 'desirability and definition of child sexual abuse material is quite properly well settled' 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 legal, but US legal parameters do not, issue which is likely to become more pressing with technological change. Douek, E (2020) The rise of content cartels: using transparency and accountability in industrywide content removal decisions. New York City: Knight First Amendment Institute, Columbia University

31 Canadian Centre for Child Protection (2019) How we are failing children: changing the paradigm. Winnipeg: C3P

32 WeProtect data generated by Crisp Consulting

33 Putnam, L (2022) Facebook Has a Child Predation Problem. New York City: Wired. Article published 13 March 2022

34 Based on subsequent discussions with Prof Lara Putnam at the University of Pittsburgh

35 Anonymous whistleblower disclosures made to the Securities and Exchange Commission

36 Based on analysis conducted by Professor Lara Putnam on 9 April 2022 and posted to Twitter

Given the clearly egregious nature of such material, and its direct contribution to driving illegal activity, we recommend that the scope of the illegal safety duty is amended, granting the regulator powers to require that companies address contextual CSA within a single coherent upstream framework.

Alternatively, the Government could signal that contextual CSA will be designated a primary priority harm, with Ofcom able to set out appropriate regulatory measures in its subsequent codes of practice.³⁷

Child abuse risks in private messaging and groups

We strongly welcome the Government's decision to include both public and private messaging in the scope of the Bill. The Online Safety Bill will not succeed unless its scope includes product features and design choices that pose the greatest risk for children.

While we recognise the need for the risks of private messaging to be addressed in a proportionate way, with appropriate safeguards in place, the Bill contains new restrictions on Ofcom's ability to require companies to use technologies to detect and disrupt grooming and child abuse.

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 nearly three quarters (74%) of cases, this contact initially takes place by private message.³⁸

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.³⁹

We have also seen evidence to suggest that private and secret Facebook groups are being used for organising and the distribution of child sexual abuse material.⁴⁰

Schedule 4 of the Bill would prevent the regulator from building into its codes of practice *any* requirement to use proactive technology to detect abuse in private messages. This would likely restrict Ofcom from being able to include in codes of practice tools already widely

used by most online services. Our working assumption is that this would prevent the inclusion of hash matching technology (including industry-standard products such as Photo DNA used to detect known child abuse); visual and text based classifiers that are widely used to detect grooming and newly produced images (including self-generated images); and metadata analysis to detect grooming and other forms of illegal behaviour.

In effect, this could significantly restrict the regulator's ability to draw upon many of the standard approaches already used by companies. It raises significant questions about Ofcom could realistically produce a Code of Practice that is capable of responding to the nature and extent of the online child abuse threat.

We recognise that Ofcom has been given stronger powers to require companies to use automated technologies to detect child abuse content, on both public and private parts of its service, through the use of a 'CSEA warning notice.' This is an important means to address high risk design features, and to provide backstop powers which will enable the regulatory regime to become more future-proof.

However, the regulator would only be able to use these powers where it assesses that child abuse is already prevalent – in other words, where significant harm has already occurred. This seems to run entirely contrary to the proactive and upstream emphasis on harm reduction proposed elsewhere in the legislation.

It also raises significant questions about the efficacy of requiring Ofcom to potentially take action against a large number of sites through the CSEA warning notice route, primarily because it is unable to do so at an earlier and more optimal point in the regime.

If the regulator is unable to sufficiently proactively tackle online grooming, the impact will be disproportionately felt by girls. NSPCC data shows that an overwhelming majority of grooming offences target girls, who are victims in 83% of sexual communication with a child offences (where this data is recorded.) Data suggests that girls aged 12-15 are most likely to be victims of online grooming.⁴¹

37 However, the bifurcated approach to harmful safety duties presents challenges in implementing this approach

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

39 figures from the National Center for Missing and Exploited Children

40 See above

41 NSPCC Freedom of Information request to police forces in England and Wales, August 2021

Tackling high-risk design choices

We welcome the regulator being given strengthened and simplified powers to respond to high-risk design choices, through being able to compel a regulated service to use approved automated technologies to detect child abuse content.

Companies may look to press ahead with a number of high-risk product choices before regulation comes into force: for example, Meta is proposing to rollout of end-to-end encryption across its Messenger and Instagram messaging and video chat products but has yet to commit to adequate child safety mitigations.⁴²

Separately, Twitter is actively developing a decentralised operating standard, which would effectively 'engineer away' the ability to perform content moderation (and in turn comply with much of the regulatory regime.)⁴³

We support of simplified CSEA warning notices in such instances, where there is a reasonable assessment that the associated risk profile has been insufficiently addressed. These represent a highly targeted, evidence-based, and proportionate response to significant safeguarding issues, but also provide a process through which the impact on a range of fundamental rights can be appropriately balanced.

Tackling online abuse on aural platforms

Tech companies are actively developing in new aural communication-based features, which are likely to be a continued area of growth and innovation.⁴⁴ The Bill legitimately excludes one-to-one aural communications, such as phone calls or FaceTime, from any form of proactive monitoring requirements. This is the right balance to protect user privacy and it should be retained.

However, we would encourage companies to be required to consider how one-to-one aural communications, where these are offered within a broader set of functionalities, could contribute towards increased child abuse risks.

We therefore recommend that companies are required to consider such services as part of their risk assessment duties (while ensuring the functions themselves are otherwise out of scope.) Unless companies are required to consider the interplay of all design choices and functionalities they provide, this presents an inevitable risk that potential grooming pathways may be missed – and that abuse vectors may shift towards parts of embedded services that are considered more readily exploitable for the purposes of abuse.

⁴² In March 2022, Meta published a White Paper setting out a range of mitigations that the NSPCC considers to be wholly ineffective, for example relying on larger numbers of children to report their own grooming. Meta (2022) Meta's approach to safer private messaging on Messenger and Instagram Direct Messaging. Menlo Park, CA: Meta

⁴³ Twitter has established its Blue Sky division to develop a decentralised social network standard. The unit is headed up by a cryptography specialist.

⁴⁴ Spectrum Labs (2022) The increasing use of audio: moderating audio and voice on your platform, a white paper.

Test three: tackling legal but harmful content

The Online Safety Bill must tackle clearly inappropriate and potentially harmful content. At present, children are regularly exposed to

- harmful age-inappropriate content, including pornography;
- targeted harassment and bullying;
- the distribution of intimate images that do not meet the threshold to be considered a child abuse image, but which still presents considerable cause for distress; and
- material that promotes or glorifies suicide and self-harm, which most major sites prohibit but often fail to moderate effectively.

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 threats may rapidly increase.

The Bill aims to offer a higher standard protection to children than adults, and in some areas has been significantly bolstered. For example, the Bill now includes duties on commercial pornography platforms to prevent children from using their services.

However, substantive questions remain about how effectively the Bill will offer universal protection to children. We are particularly concerned about the children's access assessment, which imposes a higher threshold than the ICO's Children's Code in respect of whether a service is likely to be accessed by a child.

The resulting 'child use test' may result in lower standards of protection, with a number of problematic services such as OnlyFans and Telegram likely out of scope, and the risk that harmful content is not tackled but rather displaced to other sites.

Achieving a higher standard of protection for children

The Bill requires all regulated services likely to be accessed by children to take proportionate measures to prevent them being exposed to harmful content.

The Government will set out a list of primary legal but harmful risks in secondary legislation. While the Bill sets out a this of priority criminal content, both primary priority content and priority content for children will be introduced at a later stage.

As the Carnegie UK Trust rightly notes these areas are critical for user groups seeking to understand if and how Bill will protect them in future, as well as the companies that might have to manage them.⁴⁵ As the Government has decided to list harms in the Bill, we recommend that proposed categories of primary priority content harmful to children should be set out as soon as possible.

Platforms likely to be accessed by children will have to conduct regular child safety risk assessments, use proportionate systems and processes to prevent children's exposure to harmful content, and 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 tightened systemic approach, companies will be expected to take account of the harms identified in Ofcom's risk profile, and comply with measures set out in codes of practice.

Companies will only have to mitigate risks that either designated as primary priority content, priority content or have been identified in the most recent child risk assessment. Where a regulated service identifies new forms of harmful content (non-designated content), these must be addressed through the risk assessment. The company must also notify Ofcom such that, as appropriate, it can reflect this in any future refresh of its risk profile. We continue to assert that companies should be required to tackle reasonably foreseeable harms, not just those captured in their most recent risk assessments. Furthermore, regulated sites should be expected to update their risk assessments regularly (not simply 'keep them up to date'); and Ofcom will need to regularly refresh their own risk profiles, to ensure these reflect new and highly agile harm dynamics.

In order to ensure Ofcom has corresponding agility to understand rapidly shifting risk factors, there is an appreciable need for additional early warning capacity to be built into the regime, including through the provision of user advocacy mechanisms with the capacity to identify new and emerging harms.

⁴⁵ Perrin, W et al (2022) The Online Safety Bill: Our Initial Analysis. Dunfermline: Carnegie UK Trust

Differential protections and the children's access assessment

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 in conjunction with the Bill's failure to adequately tackle cross-platform risks, is likely to result in significant amounts of harmful content simply being displaced to other sites.

Clause 31 requires companies to conduct a 'children's access assessment', and in order for a service to be in scope, that a 'child user condition' must be met. Under the clause, a service is only considered as being 'likely to be accessed by children' if there are a significant number of children who use it, or the service or any part of it 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 Bill's drafting could result in highly problematic services 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.

Although the final Bill gives Ofcom the power to determine that a service is likely to be accessed by children, there is a clear moral hazard for platforms to wait until they are 'picked off' one by one by the regulator, with platforms being able to offset any enforcement penalties against the financial commercial benefits of being able to postpone compliance.

There are legitimate questions about whether Ofcom will have the bandwidth and resources to proceed with multiple cases, or to do so quickly.

New platforms can often grow their user base rapidly and in some cases exponentially. We therefore have concerns that the children's access assessment must only be undertaken at least every 12 months. This presents a legitimate risk that with the 'child use test' in place, the Bill's drafting will fail to capture harms on the fastest-growing platforms or those that reposition their user base, and there is at best an in-built lag in the regime.

Delivering appropriate responses to harm

Under the Bill, content will be considered harmful to children where it is designated as primary priority content; priority content; or is likely to cause material risk of significant harm to an appreciable number of UK children.

These thresholds are comparatively higher than those set out in the Video Sharing Platform Regulations, in respect of protecting children from any content that might impair their mental, physical or emotional development.

Similarly, we have concerns that the quantitative threshold of an 'appreciable' number of children could result in the Bill offering weaker protections to small groups of children, and to children who may be subject to significant harm but where it is challenging, time-consuming or there are methodological barriers to being able to quantify it.

This could potentially affect children who experience targeted abuse, for example to exploit a physical or mental disability; children with particular vulnerabilities; LGBTQ+ children; and children in child abuse image sequences which do not meet the criminal threshold, but whose images are circulated at scale and who can experience pronounced trauma and re-victimisation.⁴⁶

Developing effective age assurance

The legislation will require companies in scope to introduce age assurance technologies, in order to determine whether a user is a child and therefore requires the additional regulatory protections set out in the regime.

The Bill envisages that age assurance will be a primary means for companies to discharge their child safety duties. Clause 11(3) sets out that companies will be required to use age assurance techniques that are proportionate to the nature of harm, which in respect of certain forms of primary priority content may necessitate age verification, but for other harms may be achieved through other less invasive forms of age assurance.

Although we understand Ofcom will have powers to set guidance on age assurance measures, it remains unclear whether such guidance will be legally binding, and what standards and thresholds are likely to apply.

There remains considerable merit in the Bill requiring of harm to produce a Code of Practice on age assurance technologies,⁴⁷ and in Ofcom and Government providing further information during parliamentary passage on how it envisages age assurance being used to deliver its legislative and regulatory objectives.

⁴⁶ For example, a survey of child abuse survivors conducted by the Canadian Centre for Child Protection found that a significant minority of respondents reported having been identified by someone who had seen images or videos of their abuse, and nearly 70% of respondents believe worried about this happening to them.

⁴⁷ the NSPCC supports Baroness Kidron's Private Members Bill on age assurance standards

Children's access to pornography

We strongly welcome the Government's decision to introduce new duties in relation to pornography.

Part five of the Bill places a duty on regulated services to prevent children from being able commercial on their platforms. Access to age-inappropriate pornography is a substantive concern: recent research is found that 62% of 11-13 year-olds who reported having seen pornography described their viewing as mostly unintentional.⁴⁸ Many children report that stumbling across sexual content accidentally as a distressing experience.⁴⁹

Pornographic material can distort children's views of sex, consent, and relationships, and in the context of the recent Ofsted report into 'Everyone's Invited', is a likely contributory factor to broader issues of harassment against girls.

It is important to note that part five only captures commercial pornography, with user generated material likely to be addressed separately. This presents a further complication: if user generated material is to be addressed through the child safety duties, presumably through being designated as a priority harm, a number of user-generated platforms could remain out of scope. This includes a number of high-profile user-generated adult sites, most notably OnlyFans⁵⁰ and Just for Fans.

User empowerment duties

The Bill sets out a range of measures that Category 1 companies must take to empower adult users. Specifically, these companies will be required regulated sites to provide adult users with the means to filter out non-verified users; prevent anonymous accounts from interacting with any content they generate, upload or share; and reduce the likelihood that they will encounter material posted by anonymous accounts (clause 14.)

These measures represent a targeted and proportionate response to the risk of abuse and harassment from anonymous accounts, while preserving the right to anonymity for the broad range of users (including children) who benefit from it.

However, the decision to only offer these measures to adults means that children will have reduced means to exercise choice over their online experience than adults; customise their user experience as they see fit; and to benefit from equivalent protections from anonymous abuse.

This seems wholly incompatible with the legislative objectives that children should receive a higher standard of protection than adults. The Bill should be amended to provide children with at least the equivalent means to prevent unwanted engagement with content posted by anonymous accounts from adults.

48 BBFC (2020) 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 18 tells us about the necessity for potential efficacy of emerging legislative approaches. Policy and Internet, pp1-18

49 BBFC research

50 OnlyFans will be required to introduce age verification measures to comply with the Video Sharing Platforms (VSP) regime, although it is anticipated that the online safety regime will supersede these arrangements. As a result, there is a potentially perverse outcome whereby OnlyFans is subject to a less onerous regulatory regime when the Online Safety Act takes effect.

Test four: transparency and investigation powers

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

The Bill proposes to give Ofcom an effective suite of investigation powers, although it will be crucial that Ofcom has the resources it needs to investigate how and whether platforms are complying with their safety and risk assessment duties.

We remain disappointed that the Bill fails to include information disclosure duties on regulated companies. Disclosure duties could play a valuable role in hardwiring safety duties into corporate decision-making. The Bill could usefully be strengthened by integrating this aspect of regulatory design into the proposed approach, particularly given how effectively this works in other regulated markets.

Information gathering powers

Ofcom will benefit from comprehensive information gathering powers, including the power to issue information notices for exercising, or deciding whether to exercise, relevant online safety functions. These powers cover regulated firms, web hosting infrastructure, and appear to extend to former employees of regulated companies and relevant ancillary bodies,⁵¹ which could include app stores or third parties that support platforms to discharge their regulatory duties.⁵²

The regulator will be able to launch investigations with a range of powers at its disposal. This includes the ability to commission a Skilled Persons report. Ofcom will have strengthened powers to commission a review and appoint an independent 'skilled person' to conduct it, with the regulated party being liable for the costs involved. Regulated services are required to give the skilled person assistance, which is an 'enforceable requirement' under the regime.

The regulator will also have the power to interview staff; and powers of entry and inspection.

Transparency reports

The Bill makes provision for a duty on regulated companies to prepare annual transparency reports. (Clause 64).

Transparency reports will prove beneficial if they provide meaningful and interrogable information to all interested parties, compared to existing voluntary approaches that have been widely dismissed as a form of 'transparency theatre'.⁵³ The final Bill gives Ofcom greater flexibility to decide the scope of reporting requirements on regulated firms, moving away from a closed list to the power to specify any relevant matters (schedule 8).

While the statutory provisions appear generally sound, there are some substantive questions about how the transparency regime will be enacted. Companies will only be required to publish transparency reports if they are designated as Category 1, 2A or 2B providers. However, the designation of platforms will only be made after Royal Assent, when Ofcom establishes a register of companies (clause 81). As such, we still do not know which providers will fall within these categories, and which may be excluded altogether.

The Bill's risk assessment sets out modest 10-year transparency by costs of £6.3 million.⁵⁴ If indicative of Ofcom's likely requirements, this suggests a relatively limited set of transparency measures may actually be sought.

The Bill places a duty on providers to ensure its transparency reports are accurate and complete. Although this is an enforceable condition, Ofcom would have to investigate where it has concerns about the reliability of company data, and it may not be readily obvious that erroneous or misleading information was submitted.

We therefore see merit in companies being subject to some form of quality assurance activity led by the regulator. This would 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.

51 Category 1 companies are anticipated to be the largest user-to-user sites, category 2 comprises search functions, and category 2b will capture smaller platforms (although will have both a lower and upper qualifying threshold.)

52 In doing so, mitigating concerns about the so-called 'transparency deficit' in ancillary arrangements such as GIFCT. In February 2022, a BBC investigation alleged that OnlyFans had sought to blacklist the accounts of adult performers on competitor sites by placing them on a database of terrorist material facilitated by GIFCT.

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

54 UK Government (2022) Online Safety Bill Impact Assessment. London: UK Government

Proactive information disclosure duties

We are disappointed the Bill does not introduce broad workable information disclosure duties on regulated companies.

Category 1 services should face regulatory duties to proactively disclose information to the regulator about which it could reasonably expect to be informed about.⁵⁵ For example, companies should notify Ofcom about significant changes to their products or services, or to their moderation arrangements, where these may impact upon the child abuse threat and its response to it.

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

All regulated companies should also be subject to 'red flag' disclosure requirements, in which they would be required to notify Ofcom of any significant lapses in, or changes to, systems and processes that compromise their discharge of the safety duties.⁵⁶ For example, if regulation was already in place, Meta might reasonably have been expected to report on the opaque technology issues which caused it to detect significantly less child abuse content during the second half of 2020/21.⁵⁷

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, as a valuable means of hardwiring regulatory compliance into the operation and decision-making structures of regulated companies.

⁵⁵ Principal 11 of the financial services regime

⁵⁶ For example, financial services companies are required to make reporting disclosures under the anti-money-laundering and financial services regimes, and gambling firms must report breaches against self-exclusion protocols

⁵⁷ According to Facebook transparency reports, technical problems resulted in the volumes of child abuse content being actioned by the site falling by half during this period. Facebook provided limited information about the reasons behind this considerable drop off, which in turn would result in significant declines in actionable intelligence being provided to police

Test five: enforcement powers

If online harms regulation is to succeed, the regulator must have suitably broad enforcement powers which are able to hold non-compliant sites to account and effectively change culture within the regulated companies.

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

We have substantive concerns that the proposed enforcement approach set out in the final Bill does not go far enough to incentivise compliance, nor deliver much-needed cultural change.

Senior management liability

The Bill's approach to senior management liability represents 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.

We welcome senior manager liability being strengthened compared to the draft legislation, with senior managers subject to a number of offences where they fail to comply with an information request, or knowingly seek to mislead. Criminal sanctions will now also be introduced once the regime takes effect.

However, we remain concerned this approach remains poorly targeted towards delivering child safety outcomes: senior managers will not be liable for substantive product or safety decisions, and the Government has rejected the Joint Committee's recommendation that each company appoint a 'Safety Controller', at or reporting to Board level, who would be criminally liable for serious regulatory breaches.⁵⁸ As a result, there is no direct relationship in the Bill between senior management liability and the discharge by a platform of its illegal and child content safety duties.

Under the Bill, a regulated service could demonstrate wholly negligent behaviour that exposes children to preventable child sexual abuse (and that could have been prevented by a senior manager's decision making), but the relevant senior manager would face no personal liability so long as they cooperate with the regulator.

Based on the experience of other regulated sectors – principally financial services – there is a compelling case for both corporate and senior management liability in respect of the illegal content safety duties.⁵⁹ 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, Ofcom could bring proceedings against senior managers that breach their illegal content and child safety duties, with proportionate sanctions such as fines, disbarment, or censure.

The clear deterrence value of such an approach, and the potential for adverse individual and corporate reputational effects, are obvious.

For the most significant failings, there should be provision for criminal sanctions, but only where there is clear evidence of repeated and systemic failings that result in a significant risk of exposure to 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, with the debate arguably being steered by tech exceptionalism rather than an emphasis on building a regulatory regime that draws from successful compliance in other regulated sectors.⁶⁰

⁵⁸ Joint Committee on the Draft Online Safety Bill (2021) Report of Session 2021-22

⁵⁹ Based on extensive discussions with the Financial Conduct Authority and regulated persons in the financial services regime. See also Chiu, I (2016) Regulatory Duties for directors in the financial services sector, and directed in company law – bifurcation and interfaces. *Journal of Business Law*, 2016

⁶⁰ For example, much of the tech industry has referred to the Bill's criminal provisions as 'hostage taking laws' and made spurious comparisons to authoritarian regimes

In its final response to the White Paper, the Government set up concerns expressed by tech companies about the 'potential negative impact on the attractiveness of the UK tech sector'.⁶¹

More recently, tech lobbying has sought to draw a direct relationship between personal accountability and negative impacts on free expression. The tech lobby has claimed that the introduction of senior manager liability would incentivise companies to takedown excessive amounts of content.⁶² This seems to reinforce that senior management liability is where, if built into the regime, Ofcom could expect to achieve greatest regulatory 'bite'.⁶³

As it stands, the Online Safety Bill is now weaker in this regard than the General Online Safety Bill currently being scrutinised by the Oireachtas. Ireland's legislation includes senior manager liability for both regulatory breaches and a failure to cooperate with investigations.

Confirmation decisions

The Bill makes provision for Ofcom to issue 'confirmation notices' to regulated companies. Confirmation notices will allow the regulator to direct companies to take specific steps to comply with regulatory requirements, and to remedy one or more areas where a platform is non-compliant.

While the approach to confirmation notices is generally sound, clause 116 introduces restrictions on Ofcom's ability to require a company to use proactive technology to identify or disrupt abuse in private messaging. These restrictions strike the wrong balance between protecting user privacy and promoting the safety and privacy of children at risk of sexual abuse.

While we recognise that additional safeguards may be desirable before the regulator could instruct a platform to use certain proactive technologies, for example more complex types of classifiers, this clause would prevent Ofcom from being able to require the use of any proactive technology in private spaces.

This would effectively prohibit the regulator from being able to recommend solutions that are already widely deployed across the industry, including the use of hash scanning technology used to identify and remove known child abuse images.

It would be clearly disproportionate to restrict Ofcom's ability to require the adoption of industry-standard techniques, most notably PhotoDNA and CSAI Match, and risks building an unhelpful narrative that well-established, privacy-preserving techniques are somehow problematic.

Effective financial penalties

The Bill contains steep financial penalties for firms that breach their regulatory obligations. Regulated companies could face fines of £18 million or ten per cent of turnover (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 engagement approach, it is questionable whether they offer sufficient deterrent value for the largest tech companies. For companies with significant 'cash in hand', the microeconomic effect of fines will be blunted, and they are likely to have modest impact at best on senior management behaviour and corporate decision-making.⁶⁴

Business disruption measures

The legislation imposes a range of ambitious service and 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 groups.

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 and web hosting services being in scope.

Decisions by Visa, MasterCard and Discover to stop processing transactions on Pornhub resulted in the service making significant changes to how it moderates user generated content. Similarly, decisions by Amazon Web Services (AWS) and Cloudflare to suspend hosting of a number of sites, including in response to concern about the adequacy of their child abuse and content moderation policies, has been instrumental in making sites adopt safer approaches or disrupting potential harms.⁶⁵

61 UK Government (2020) Response to Online Safety Bill White Paper. London: UK Government

62 See for example Google's response to the Joint Committee

63 It also actively conflates the systems based approach of this regime with one focussing on content.

64 Investigations and appeals can be lengthy, and by the 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

65 Rosenzweig, P (2022) Countering harmful content: a research agenda. Lawfare blog. Posted 11/02/22

Test six: user advocacy arrangements

Effective user advocacy is integral to the success of the regulatory regime. During parliamentary passage there is an important opportunity to embed effective user advocacy into regulatory design to ensure online safety regulation delivers better outcomes for children. Well-designed user advocacy arrangements should be put in place to ensure the regulated settlement is not disproportionately skewed towards the interests of industry, rather than children.

We encourage the Government to set out much more ambitious proposals – and deliver user advocacy mechanisms that secure the confidence and support of victims' groups.

Creating a statutory user advocate for children

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.

Fully-fledged statutory user advocacy arrangements are used in nearly all regulated consumer sectors including energy, water, post, and transport. They play a key role in representing users, particularly vulnerable groups, and ensuring that their voices are heard and appropriately counterbalanced against the backdrop of well-resourced and vocal regulated companies.⁶⁶

In respect of the Online Safety Bill, equivalent provisions have not been introduced. Instead, Ofcom will be required to establish its own arrangements to understand the interests and experiences of service users. Ofcom will be required to publish statements about any research or consultation that it undertakes.

User advocates' insights help both regulators and regulated companies make better decisions and drive better, more user-focussed outcomes.

Unless user advocacy arrangements are built into the regime, children who have been or are at risk of sexual abuse will receive less statutory user advocacy protections than users of a post office or passengers on a bus. It is vital that children have an advocate to reflect their needs, and user advocacy mechanisms are crucial to ensure the regime can deliver on its stated goal of delivering a higher overall standard of protection for children.

Without a user advocate for children, we risk create a regime where children's voices are lost, and complex safeguarding issues do not receive the attention they need in the new regulatory regime. A user advocate for children can use its expertise and insight to ensure that children's needs are effectively met – and can serve as a powerful, consistent and well-resourced voice to cut through and counterbalance industry interventions.

User advocacy means better regulatory outcomes

User advocacy mechanisms are a crucial part of the regulatory regime. Put simply, 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 offer support to the regulator to understand often complex child abuse issues; and that is able to 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. If there is an inappropriately scaled, poorly focused or insufficiently resourced response, this is likely to significantly weaken the regulator's ability and appetite to deliver meaningful outcomes for children.

User advocacy must be seen as an integral part of delivering a regulatory settlement that puts children first, and that responds to the lessons of regulatory asymmetries in other markets. A fully fledged user advocacy mechanism can:

- act as an early warning function that strengthens the overall regime: the Bill sets out a systemic risk assessment process that is underpinned on and reliant on early identification of new and emerging harms. Child online harms are characterised by their highly agile and constantly evolving nature, and an early warning function to flag new and emerging threats, as can be delivered most effectively through a user advocacy mechanism, is therefore vital;
- *provide much needed counterbalance to industry*: 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.

⁶⁶ The value of levy 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

In most other regulated markets, these risks are addressed through strong, independent advocacy models. 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.

- *prevent the risk of an 'evidence trap':* with so much of the regime being left to Ofcom to establish through a framework of codes and guidance, there is a palpable need to address the risks that tech interests seek to skew the evidence base upon which Ofcom will correctly base its decisions.

There is a pronounced risk that without an effective counterbalance, large tech companies will attempt a concerted attempt to capture independent and expert voices; commission, fund or enable highly selective research with the intent to skew the evidence base; and then challenge any decisions which run contrary to the evidence base it has created.

These tactics are not new – we have previously seen similar tactics used by other regulated sectors, such as the tobacco industry.⁶⁷ In recent years, we've seen tech companies look to pursue similar tactics, including attempts to frustrate evidence on the nature of AI risks;⁶⁸ and through granting privileged access to data sets for favoured researchers.⁶⁹

But if they are not proactively addressed, they could represent a significant challenge to the regime's overall effectiveness.

- *Ensure complex safeguarding issues are effectively built-in to the regime:* user advocacy is essential to drive Ofcom to prioritise children's issues and ensure regulators have an accurate, well informed understanding of new and emerging issues. In a rapidly changing sector, Ofcom will need to be equipped with a robust and agile understanding of harm; feel confident in understanding the interplay between technological and market change, children's use of products and resulting safeguarding risks; and feel comfortable confident that it has a robust understanding of systemic issues that will likely require their attention.

Ofcom's CEO has recognised that, in respect of the Online Safety Bill, independent expertise will be more important to the discharge of their functions than in any other part of its regulatory remit.⁷⁰

However, without well-established user advocacy mechanisms in place, there are legitimate questions about how Ofcom can benefit from the level of specialist expertise, evidence and critical challenge that is likely to be required.

How should user advocacy be funded?

The industry levy is a highly appropriate mechanism for funding statutory user advocacy for children. This is entirely consistent with the well-established 'polluter pays' principle and corresponds to the funding arrangements for user advocacy in other markets.

A levy model is wholly proportionate and reasonable when considered against the commercial return available to companies that offer their services to children but fail to protect them from reasonably foreseeable harms.

NSPCC analysis suggests the average cost of user advocacy provision in comparable markets is £4.1 million per year.⁷¹ User advocacy is an exchequer-neutral policy, and it represents only a minimal additional burden on regulated firms (the 10-year total costs of levy fees is estimated at £313 million.)⁷²

However, it is also reasonable to issue that well-established user advocacy mechanisms could actively contribute towards the delivery of more effective regulatory outcomes, and in turn, the functioning of a safer and more compliant set of industry approaches. User could therefore deliver broader social, economic and societal benefits, and bolster the case for online harms regulation in strictly economic terms.

67 For example, see Abdalla, A; Abdalla A. (2021) The Grey Hoodie Project: Big Tobacco, Big Tech, and the Threat on Academic Integrity. Proceedings of the 2021 AAAI/ACM Conference on, Ethics and Society. Toronto: University of Toronto; Cambridge, MA: Harvard Medical School

68 For example, the high profile case of Timnit Gebru, in which she was asked to withdraw a research paper on algorithmic bias by her employer, Google

69 The Research Director of the Shorenstein Center on Media Politics and Public Policy At Harvard, Joan Donovan, has voiced that 'it's frustrating to see an effort Facebook has put into academic capture over the last four years, selecting certain firms to receive special datasets [...] This is the playbook from Big Tobacco and Big Oil.' Comments posted to Twitter, January 2021

70 Comments made by Dame Melanie Dawes in her oral evidence session to the Joint Committee on the Draft Online Safety Bill

71 Forthcoming analysis

72 HM Government (2022) Online Safety Bill Risk Assessment. London: HM Government




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	Overall
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	Overall
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 **Overall**





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 **Overall**




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 **Overall**




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 **Overall**



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 experienced 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](https://www.nspcc.org.uk)