NSPCC

How to win the Wild West Web

Six tests for delivering the Online Harms Bill

September 2020

EVERY CHILDHOOD IS WORTH FIGHTING FOR

Contents

Summary

In the coming weeks, the Government will decide on legislation that could finally protect children from online abuse. If it acts with urgency and ambition, it can secure an Online Harms Bill that delivers tough but proportionate regulation, and that sets a global standard.

But if the measures fall short, children will continue to face avoidable harm. One in five UK internet users[1] will face online abuse that continues to increase in both scale and complexity. The cost of industry inaction will continue to be felt by children, families and society.[2]

After a year in which they have faced unprecedented online risks, fuelled by the public health emergency but driven by the long-term failure of self-regulation, it couldn't be clearer that children deserve better than the status quo.

Earlier this year, the Prime Minister told participants at his Hidden Harms Summit he was determined to take tough but necessary action to hold social media companies to account. He heard the words of a mother whose 12-year-old daughter Freya was subject to online abuse: "Our children should be safe in their bedrooms, but they're not. They should be safe from messages from strangers if their accounts are on private, but they're not."

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

This report reaffirms the case for action – but it is clear the Government will only deliver on its ambition to make Britain the safest place in the world online[3] if it is bold and ambitious in its plans. If regulation is poorly designed, or the regulator isn't given the powers it needs, children will continue to face otherwise preventable harm.

Last year, in conjunction with Herbert Smith Freehills, the NSPCC published clear proposals for a regulatory model.[4] In this report, we set out a series of tests that the Online Harms Bill must meet if it is to deliver for children, and against which the Government's commitments should be judged.

If it meets each of these tests, the result will be a highly effective regulatory regime, and a Duty of Care that gives children long overdue online protections.

Six tests for the Online Harms Bill

An expansive, principles-based Duty of Care

Statutory regulation must be tough but proportionate, and it should deliver the strongest possible protections from abuse for children. This means the Duty of Care must be realised through a principles-based approach which is broad, future proof and that applies expansively.

In the event that harm occurs, a platform would breach its Duty of Care if it failed to demonstrate sufficiently rigorous processes to identify or mitigate reasonably foreseeable harm, or if children had been put at material risk as a result of systemic failures that could reasonably have been addressed.

The Government must resist calls for a more prescriptive, and by implication less ambitious, approach. It is precisely because the Duty of Care requires platforms to assess the risks on their own sites, not just to follow a tick box set of remedies, that regulatory requirements will be hardwired into platform decision making – and that significant cultural change will be achieved.

Tackling online child abuse

The regulator must demonstrate an ambitious and determined focus on tackling online child abuse.

Ofcom will rightly be judged on how effectively it disrupts both online grooming and the production and distribution of child abuse images. It must prove capable of responding to constantly evolving abuse and highly agile threats.

Despite the understanding that tackling child abuse requires an emphasis on only illegal material, there are significant problems with abusive images that may not meet the criminal threshold, but which have significant potential to cause harm, signpost to illegal material, or re-victimise the children involved. More proactive processes to respond to such images will be required, which should include consistent takedown processes.

Platforms should have a duty to collaborate on child abuse risks, and should be subject to enhanced regulatory measures for high-risk design features that increase the risk of technology-facilitated abuse, including livestreaming, private messaging and end-to-end encryption.

1 Data from the Information Commissioner's Office.
2 The Center for Humane Technology maintains a Ledger of Harms that lists the 'negative impacts of social media that do not show up on the balance sheet of companies, but on the balance sheets of society.'
3 Department for Digital, Culture, Media and Sport (2017) Internet Safety Strategy Green Paper. London: DCMS.
4 NSPCC (2019) Taming the Wild West Web. London: NSPCC.

Tackling legal but harmful content

In its interim response to the white paper, the Government set out differentiated expectations for illegal content, and that which is legal but causes harm.[5] This effectively requires platforms to only adopt clear policies on legal but harmful material, and enforce them effectively.

The regulator must adopt a child-centred and harm-based approach to legal but harmful content. Its regulatory approach must take decisions that are appropriately balanced against freedom of expression, but that respond to the very significant potential for harm that comes from platform mechanisms that promote or algorithmically suggest harmful content, including suicide, self-harm content and preparatory child abuse images. These clearly require an effective regulatory response, and the Government has positive obligations to protect children online.

Child users should receive protection that is proportionate to the likely harm caused. The Government must therefore ensure that any differentiated Duty of Care does not result in companies facing a perverse incentive to adopt weaker community standards, because in turn it will result in less onerous regulatory requirements.

In accordance with a risk -based approach, the regulator should signal its intention to apply enhanced regulatory scrutiny on content that is likely to be harmful to children.

Transparency and investigation powers

Comprehensive transparency powers are crucial to the regulator's success. Unless Ofcom has robust investigatory and information disclosure powers, there will be a clear information asymmetry - and this could mean it is forced to take decisions on low quality evidence, or is less inclined to propose more ambitious measures.[6]

It is not enough to rely on industry transparency reporting. Such arrangements will only be beneficial if they provide significant and interrogable information, compared to existing approaches that are widely dismissed as a form of 'transparency theatre'.[7]

Platforms should face new information disclosure duties, including a requirement to proactively disclose to the regulator any information it could reasonably expect to be informed about, and to 'red flag' cases where failings could put children at risk. To embed a safety-by-design approach, sites should be required to undertake a risk assessment if they plan to introduce new services or amend their existing ones.

Criminal and financial sanctions

If the regulator is to effectively hold platforms to account, it requires comprehensive enforcement powers that ensure companies comply with the Duty of Care. Both platforms and senior managers must be liable to financial and criminal sanctions.

The powers available to the regulator must clearly correspond to the size and scope of the companies it regulates. We support GDPR equivalent fines, but for the largest companies the deterrence value of such fines is at best unclear.

The Government must therefore commit to both corporate and senior management liability.

The Bill must introduce a Senior Managers scheme that imposes personal liability on directors whose actions consistently and significantly put children at risk. For the most serious of failings, the threat of personal prosecution should apply.

Industry groups have fiercely opposed personal liability, but the case for criminal sanctions in providing incentives to take action is compelling.[8]

User advocacy arrangements

As part of the regulatory settlement, it is essential there are effective arrangements in place for civil society to represent children's interests in regulatory debates. It will be necessary for civil society to support the regulator in understanding often complex child abuse risks; provide high-quality evidence of a sufficient regulatory threshold; and to demonstrate areas of concern or non-compliance.

5 Department for Digital, Culture, Media and Sport (2019) Online Harms White Paper. London: DCMS.

6 This is likely to be particularly apparent in respect of ex ante measures. Beverton-Palmer, M et al (2020) Online harms: bring in the auditors. London: Tony Blair Institute for Global Change.

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

8 This is perhaps best expressed by Twitter's CEO Jack Dorsey. Asked why his payments company Square, of which he is also the CEO, seems to operate more smoothly he said: 'We had to get every single thing right. There's a lot of regulation around payments. If you do something wrong, you go to jail.' Comments made in a January 2019 interview with Rolling Stone https://www.rollingstone.com/culture/culture-features/twitter-ceo-jack-dorsey-rolling-stone-interview-782298/

Perhaps most crucially, the regulator is unlikely to deliver the strongest possible outcomes for children unless there is a strong civil society counterbalance to well-resourced industry interventions.

This is particularly important given that some companies might seek to frustrate or delay the regulator's work, and the heavily limited potential for children to exercise the redress options that the Online Harms White Paper proposes for adult users.

In order to create a 'level playing field' for child users, and secure the regulator's focus on child abuse risks, the Government should commit to statutory user advocacy arrangements for children, funded by the industry levy. This mirrors established user advocacy arrangements in many other regulated sectors, reflects the urgency of the child abuse threat, and responds to the inherent vulnerability of children as users of internet services.[9]

The urgency of Coronavirus

The importance of the Online Harms Bill has never been clearer than during the pandemic. The magnitude of child abuse risks vividly underlines why tech firms must finally be held accountable for the harm caused by their sites.

Lockdown created a perfect storm for online abuse. We don't yet know the true scale of online abuse during the pandemic, but we do know young people spent longer on platforms with fewer moderators.[10] We also know that offenders viewed Covid-19 as an opportunity to target often vulnerable and lonely children.[11]

No one could foresee the circumstances of the pandemic, but when the perfect storm rolled in, tech firms hadn't fixed the roof. The failure to design basic child protection into their services, and invest sufficiently in technology that could disrupt abuse, meant that social networks could be exploited ruthlessly.

But the pandemic also showed this could have been different. The same platforms that have dragged their heels on child abuse for many years responded impressively to the disinformation threat, rolling out design features in days that we were previously told might not be possible in months.[12] Platforms can act quickly and comprehensively when required.

A Duty of Care is more important than ever. It will ensure that, in future, children will not have to face the risks they do today. But it will also require platforms to be ready for structural changes in the threat.[13] As a result of the pandemic, children have changed the way they socialise and learn, we've seen the mass adoption of high-risk video chat and livestreaming technology, and long-term changes to working patterns may result in higher demand for child abuse images, and an increase in grooming to fuel it.

Children have long needed comprehensive and ambitious action to keep them safe online. Recent months have created an unarguable case to deliver a Duty of Care that disrupts and prevents the full range of online harms faced by children.

It is now time for the UK Government to translate the Prime Minister's personal commitment into a world-leading model of regulation - and deliver on the ambition for Britain to lead the way in protecting children from abuse online.

9 For example, Recital 38 of the General Data Protection Regulation states that 'children merit specific protection as they may be less aware of the risks, consequences and safeguards concerned and their rights in relation to [online services].'
10 Digital, Culture, Media and Sport Select Committee (2020). Second Report of Session- Misinformation in the COVID-19 Infodemic. London House of Commons.
11 Europol (2019) Exploiting Isolation: Offenders and victims of child sexual abuse during the Covid-19 pandemic. The Hague: Europol
12 Platforms rushed out new design features to frustrate the spread of misinformation, for example WhatsApp introduced new limits on the forwarding of user messages.
13 Europol (2019) Exploiting Isolation: Offenders and victims of child sexual abuse during the Covid-19 pandemic. The Hague: Europol.

Background

Why do we need social networks to be regulated?

Covid-19 has underlined that technology is central to children's lives. Around half of UK children aged 12 have at least one social media account, despite the minimum age requirements for most sites being 13. By age 13, that figure rises to almost two-thirds.[14]

During the pandemic, social networks allowed children to stay in touch with their family and friends - and many platforms were a lifeline for children and young people. Today, social media is a ubiquitous part of childhood, and an inescapable utility.

For too long, social networks have been allowed to treat child safeguarding as an optional extra. Despite a wide range of potential harms, many platforms have considered online safety as peripheral to their business models, and they haven't invested in or prioritised keeping children safe.

As a result, we don't have the same protections in place online as offline, and children are left exposed to unacceptable but avoidable risks online.

After a decade of insufficient action, the challenge is significant, but not insurmountable. Rapidly developing technology creates new opportunities to initiate, maintain and escalate abuse. The scale and complexity of the online threat is growing.

But this can change. If the Government acts with urgency and ambition, it can deliver an Online Harms Bill that can disrupt online abuse - with tough but proportionate regulation that delivers a world-leading model for protecting children online.

What are the risks to children on social networks?

Children face a range of abuse risks online, from the production and distribution of child abuse images, to the harmful effects of exposure to inappropriate content, to the growing scale of grooming facilitated by social networks. Platforms provide new opportunities for groomers to initiate, maintain and escalate their abuse.[15]

With so many children using social networks, gaming and messaging sites, it means that today's young people are increasingly exposed to the threat of abuse, from both adults and their peers. Groomers can readily exploit the design features of social networks to target significant numbers of children, and to move them from well-known open platforms to encrypted apps and sometimes unscrupulous messaging sites.

New types of technology, notably livestreaming and video-chat sites, have provided new opportunities for abusers to control and coerce children. In a rush for market share, platforms have rapidly expanded video-chat products before appropriate safety measures can be developed and rolled out, or with deeply concerning design features in place that clearly prioritise user growth over safety.[16]

Social networks have consistently failed to address the problems on their sites – and in most cases, in the absence of either legal or commercial drivers, tech companies have failed to adequately integrate child safeguarding into either their business models or the design of their services.

Neither is it clear whether more competition will incentivise online platforms to sufficiently address the risks their systems pose to users, at least in the short term. The network effects of children being on the same platform as their friends have so far trumped concerns about safety.

14 Ofcom (2020) Children and parents: media use and attitudes report. London: Ofcom.
15 National Crime Agency (2019) National Strategic Assessment: working together to end the sexual exploitation of children online. London: National Crime Agency.
16 For example, Facebook rolled out its Messenger Rooms video-chat platform during the pandemic, in response to the rapid user growth of Zoom, which enables up to 50 participants to join a video call. Invites can be sent to anyone with an email account, regardless of whether they are a Facebook user.

On the rare occasions where there has been action, for example Instagram's commitments to tackle self-harm and suicide content following high levels of public concern, this has been largely piecemeal action by a single site, rather than a cross-platform race to the top.

The Competition and Markets Authority's vision of safety as a measure of a well-functioning market is a distant prospect.[17]

The extent of technology-facilitated abuse

For children subjected to technology-facilitated abuse, the impacts can be life-changing. Despite the common misconception that online abuse is less impactful, NSPCC research has shown that the impact of 'online' and 'offline' abuse is the same, no matter how the abuse took place.[18]

As technology has provided new ways for offenders to commit abuse, the onus has been on social networks to do everything they can to make their platforms safer. The scale and extent of online abuse demonstrates how comprehensively social networks have failed to act.

As self-regulation has failed to step up to the challenge, and more than a decade has passed since the Byron Review first called for a voluntary Code of Practice, the risks have only increased:

Online grooming on social networks has become an urgent challenge.

In England and Wales, since 2017/18 there have been over 10,000 police-recorded offences for sexual communication with a child.[19] 70 per cent of offences (where the data were recorded) took place on just three sites: Facebook, Snapchat and Instagram. Although partly a function of scale, as the largest social networks these sites have considerable resources to tackle abuse occurring on their platforms.

In 2019, the Internet Watch Foundation identified 132,600 URLs containing **child abuse imagery**, of which 46 per cent contained images of children aged ten or under.[20] The National Center for Missing and Exploited Children (NCMEC), the global clearing house for child abuse reports, processed 16.9 million such reports last year, containing 69.1 million photos, videos and files.[21]

Social networks will argue that progress has been made in the removal of child abuse images; and while this is the case, platforms have consistently failed to tackle the **production of new child abuse images**, including self-generated photos and videos.

These are often produced as a result of coercion on social networks, livestreams and video-chats. Once abuse has been photographed or filmed, or a child has been persuaded to share such images of themselves, significant and long-lasting harm has already been done.

According to NSPCC research, more than one in seven children aged 11-18 (15 per cent) have been asked to send **self-generated images and sexual messages**.[22] Seven per cent of 11-16 year olds say they have shared a naked or semi-naked image of themselves. Research shows an average of one child per primary class had been sent or shown a naked or semi-naked image online by an adult.[23]

Groomers are able to exploit the design of social networks, using algorithmically-profiled friend suggestions to infiltrate peer networks, and to establish contact with children that can rapidly escalate into coercive sexual requests.

However, platforms have often adopted a united front to frustrate or delay external action on child abuse risks. 'This prevents any individual company from receiving too much opprobrium for any particular decision',[24] or for not doing enough to protect children in the first place.

17 Competition and Markets Authority (2020) Online platforms and digital advertising market study. London: CMA.
18 Hamilton-Giachitsis, C. (2017) Everyone deserves to be happy and safe. London: NSPCC.
19 NSPCC data, sourced from Freedom of Information requests.
20 IWF (2020) Press release.
21 NCMEC (2020) 2019 reports by electronic service provider. Washington, DC: NCMEC.
22 NSPCC (2018) NetAware research on file.
23 NSPCC (2019) Children sending and receiving sexual messages: a snapshot. London: NSPCC.
24 Douek, E. (2020) The rise of content cartels: Urging transparency and accountability in industry-wide content removal decisions. New York City: Knight First Amendment Institute, Columbia University.

The impact of Coronavirus

Coronavirus has resulted in the highest risk of online child abuse that has arguably ever been seen.

While no-one could reasonably have foreseen the circumstances of the pandemic, the current crisis has shone a light on the existing weaknesses in how platforms are designed and run. Children have been exposed to unacceptable harm, not only because of the public health crisis, but the long-term failure of many platforms to invest in tackling the child abuse threat.

At the start of the pandemic, NSPCC warned of a three-fold 'perfect storm' which could result in a spike in online harms:

- Platforms were facing pressures in sustaining their moderation processes, in some cases being forced to rely on artificial intelligence (AI) that is often used to triage but not make final decisions on more complex harms, for example grooming;
- Children were spending additional time online during lockdown, with many likely to be experiencing heightened emotional distress. Surging demand for services only exacerbated the moderation pressures that social networks and gaming sites would face;
- Intelligence from Europol and the National Crime Agency (NCA) quickly warned of a significantly increased threat, including a 'surge' in child sexual abuse material. Offenders were readily able to exploit a lack of investment in proactive safety, and a legacy of consistently poor design choices. Some abusers were readily able to identify and share information on which sites were performing particularly poorly, further exacerbating the abuse threat.[25]

Those risks translated into actual harm, and while it may be some time before we know the full extent of child abuse during the pandemic, the early indicators suggest this could have been considerable.

In April 2020, the National Center for Missing and Exploited Children received over 4 million reports of online child abuse, 400 per cent the recorded rate in April 2019.[26]

The Internet Watch Foundation reported that, in the first month of lockdown, industry compliance with takedown requests dropped by 89 per cent. Simultaneously, there were over 8.8 million attempts to access child abuse imagery on three major platforms.[27]

Platforms significantly scaled back their moderation efforts, and some sharply reduced the takedown of child abuse, suicide and self-harm material. For example, Instagram removed 34 per cent fewer child abuse reports during April to June 2020, and 74 per cent fewer suicide and self-harm images, compared to the site's rolling 12-month average.[28]

Facebook's most recent reporting suggests that it may have applied a differential approach to moderation during the pandemic. Although the company correctly prioritised its resource on the most serious illegal forms of content, there are questions whether moderation resource was directed to some of its products over others – for example, it appears Facebook itself was far more able to maintain content moderation that Instagram.[29]

25 Europol (2019) Exploiting Isolation: Offenders and victims of child sexual abuse during the Covid-19 pandemic. The Hague: Europol.
26 According to US press reports.
27 Internet Watch Foundation (2020) Millions of attempts to access child sexual abuse online during lockdown. Cambridge: IWF.
28 Facebook Transparency Report, published August 2020.
29 During the same reporting period in which Instagram recorded sharp falls in content removals. Facebook increased the number of child abuse content it removed compared to the previous quarter, and it saw a considerable but less significant decline in removals suicide and self-harm content.

Test one: The Duty of Care

The regulation of online harms is challenging, but social media platforms are not beyond regulation.

As the Government proceeds towards legislation, it is vital that it delivers a well-designed, proportionate regulatory framework that delivers the strongest possible protections for children. That means the Government must realise the ambition of a principles-based approach, underpinned by a broad and future-proofed Duty of Care. We should resist the calls for a more prescriptive and by implication less ambitious approach.

Why a Duty of Care is essential

In its White Paper, the Government rightly recognised that statutory regulation is a necessary and proportionate response to the scale of online harms.

Continued self-regulation would be a wholly insufficient response to the growing and complex risks that children face – but failing to deliver an effective regulatory regime could equally expose children to otherwise avoidable harm.

In February 2019, the NSPCC and Herbert Smith Freehills published our detailed proposals for Duty of Care regulation. Drawing heavily on the excellent work undertaken by Perrin and Woods,[30] we envisaged an expansive Duty of Care that required online platforms to identify reasonably foreseeable risks caused by the design or operation of their sites – and to have a legally enforceable requirement to take reasonable measures to mitigate them.

It is vital that the Government commits to such a broad-based, overarching approach. This would require platforms to demonstrate that children's potential exposure to harms have been actively considered when making decisions, and the site's products and processes are consequently safe or low-risk by design. Compliance would not be assessed solely against a prescriptive and pre-determined set of requirements.

If implemented correctly, this is a purpose-driven and agile approach – and it will actively hardwire compliance into firms. But it will also deliver a far greater prize: the emphasis on systemic risk should bring about much needed cultural change across platforms that have previously been able to decide for themselves whether and how they protect children.[31]

The Duty of Care is desirable precisely because it is broad-based – setting out the required outcome, to prevent harm to children, but not prescriptively setting out a detailed process for how it should be implemented.

This means the strategy for reducing online harms sits with the companies subject to the regulation, that are best placed to deliver the context and platform specific responses that are required.

Regulation should be implemented according to a risk-based approach. This will enable companies to focus on substantive rather than technical compliance, and to direct their resources to tackling the most problematic of harms, for example grooming and the production and distribution of child abuse images.

Scope of a Duty of Care

It should be for the regulator itself, in consultation with civil society and industry, to develop a set of regulatory outcomes, and an outline set of harms that must be tackled.[32] This will ensure the regulator is able to bring to bear its regulatory, market and technical understanding when developing its approach.

It will also ensure that Ofcom can review and amend this list, as part of its work planning and strategic review exercises, and in response to ongoing market or technology changes, and any shifts in the threat landscape.

However, we anticipate the regulator should have statutory responsibility to tackle the most serious illegal harms, including grooming and the production and distribution of child abuse images. This approach has the benefit of providing clear direction to the regulator about the importance of tackling online child abuse – and it is consistent with the Government's commitment that protecting children should be a centrepiece of the Bill.

Any list of harms should be non-exhaustive - and there should be clear incentives for platforms to identify and protect against any emerging risks, including as a result of introducing new products or technology.

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

31 Research from DotEveryone finds that more a quarter (28 per cent) of tech sector workers have seen a decision made about technology which they feel could be damaging to society or users. 78 per cent felt they need more practical resource to enable them to think about the societal impact of their products. Miller C, Coldicutt R. (2019) People, Power and Technology: The Tech Workers' View. London: Doteveryone.

32 This reflects the risk-based model advocated by Sparrow (2011) in which the regulator should exercise choices about which harms to focus on, and using the range of instruments available to it, should prioritise those harms that most impede the delivery of regulatory outcomes. Sparrow, M. (2011) The Regulatory Craft: controlling risks, solving problems, and managing compliance. Washington DC: Brookings Institution.

Under the Duty of Care approach, firms would be required to ensure their sites are safe at a system level – this means ensuring that their products are safe or low risk by design.

In order to demonstrate compliance with the Duty of Care, a social network would need to demonstrate it had taken reasonable steps to ensure its products and processes are both designed and operated in a way that minimises the potential for harm.

We envisage the regulator should provide a list of non-exhaustive examples to guide, but not direct, industry compliance. It should use ongoing thematic reviews to test company performance, update industry guidance and best practice, and where necessary to determine the need for enforcement action. Companies must not be allowed to tie up the online harms regime in so many checks and balances that the regulator cannot do its job. Ofcom is a well-established regulator set up to take difficult decisions about powerful media through its broadly-based board and evidence focussed processes. Ofcom has proved that it can be trusted with hard decisions that balance rights and the regime should allow it to continue to operate without burdensome new process.[33]

A 'whole system' approach

Online abuse is rarely siloed on a single platform or app. It is therefore vital that the Duty of Care regulation is applied expansively, with platforms being responsible for harms that happen as a direct consequence of the design of their site or activity enabled by it, even if the victims of harmful activity may not themselves be users of it.

Sites should demonstrate they have coherent plans to contribute towards a 'whole system' approach to risk. For example, they should have plans to tackle well-established online grooming pathways, in which abusers exploit the design features of social networks to make effortless contact with children, before migrating them elsewhere; and gaming services must be ready to disrupt offenders that target children on their sites while simultaneously talking to them on ancillary chat services.[34]

The Duty of Care must be applied on a 'best endeavours' basis. While all platforms should reasonably be expected to adhere to a set of minimum safeguarding standards, the regulator should determine compliance based on the expertise and resources likely to be available to the online service.

This assessment of proportionality should inform the approach, and form part of a wider package of measures to ensure regulatory design avoids unintended consequences, including barriers to market entry.

As a minimum, the regulator should recognise that the ability of larger sites to identify reasonably foreseeable risks (including emerging harms), and to commit engineering and operational resource to address them, will be substantively greater than for smaller sites.[35]

Larger platforms could reasonably be anticipated to contribute towards a greater strategic role in addressing 'whole system' risks, and incentivised to develop new products or mechanisms to tackle them.[36] This is addressed later in the report.

Precautionary principle

The regulator should be instructed to act on a precautionary principle basis: if there is reasonable indicative evidence of harm to children, it is both appropriate and prudent to regulate the cause of it.

While there is clear evidence that social media platforms are enabling technology-facilitated abuse, the evidence base continues to develop around wider potential harms. Given their inherent reluctance to share data, many platforms arguably frustrate the development of evidence-based understanding of the harms related to their sites. As a result, it seems unlikely we will be able to fully understand the scale and extent of online harms until and unless a regulator has the information disclosure powers to compel firms to disclose data, or platforms can be incentivised to share it.[37]

Under a precautionary principle approach, firms would be incentivised to share verifiable data where this could demonstrate their products were not causing harm. In turn, this disclosure could result in a lessening of the regulatory burden, and reduce the costs of compliance.

The 'evidence gap' is particularly acute in respect of algorithmic profiling and content amplification – the potential impacts of which will probably only be reasonably understood if the regulator, and external researchers, have access to data and indeed to the algorithms directly.

33 Ofcom's proven regulatory ability is set out in Perrin and Woods' proposals, published by the Carnegie UK Trust in April 2019.
34 Helm, B (2020) Sex, lies and video games: Inside Roblox's war on porn. Published in Fast Company magazine.
35 The principle that a higher standard of care is expected from larger companies is established in case law. For example, the case of Thwaytes vs Sotheby's (2015).
36 While being mindful of Evelyn Douek's arguments that larger platforms may seek to consolidate their power or impose their corporate processes through the provision of tools made available for cross-industry use. Douek, E. (2020) The rise of content cartels: Urging transparency and accountability in industry-wide content removal decisions. New York City: Knight First Amendment Institute, Columbia University.
37 For example, the Commons Science and Technology Committee warned that a lack of data from social media companies was 'holding back the development of the evidence base.' In his oral evidence to the inquiry, Professor Andrew Przybylski of the Oxford Internet Institute, described a 'fundamental informational asymmetry' between industry teams and academic scientists.

Test two: Tackling online child abuse

The regulator must be ambitious and determined in its commitment to tackling online child abuse. Ofcom will rightly be judged in how effectively it can protect children from abuse risks that continue to grow, both in scale and complexity.

Technology-facilitated grooming

In order to tackle the risks of online grooming, platforms must be required take proactive steps to identify and disrupt illegal behaviour on their sites.

At present, groomers are all too easily able to exploit the design features of platforms to identify and make contact with large numbers of children, before migrating them to encrypted messaging and livestreaming sites, where they can rapidly escalate the process of exploitation, coercion and control.

Sites should be required to take reasonable measures to identify and prevent grooming – with a recognition that better design choices, and the proactive use of technology, can actively frustrate groomers from making initial contact with young people. In turn, this will disrupt grooming pathways at the earliest possible stage, and prevent the potential for further upstream harm.

Platforms should be required to adopt algorithms that proactively flag and identify accounts displaying suspicious patterns of behaviour. Such analysis can be conducted in a non-intrusive way, using metadata to flag accounts which should be reviewed by moderators.

Sites should be encouraged to invest in artificial intelligence, with classifiers that can detect linguistic, syntax and other situational indicators of abuse.[38] Platforms should also be required to consider the grooming risks associated with design features, and assess whether appropriate risk mitigations are in place. If the risks cannot be successfully managed, sites should consider restricting them.[39]

Social networks and gaming sites should explore options for intelligent design, for example the potential to build additional friction into the user experience of higher-risk design choices. Platforms could take a range of carefully considered steps, such a restricting the ability to send or receive direct messages, or to video-chat, for a 48-hour 'cooling off' period after a friend request is accepted.

In creating additional friction, platforms could substantially frustrate the potential for their design features to be readily exploited by abusers. This could meaningfully contribute towards how a platform delivers its Duty of Care.

Child abuse images

Online harms regulation must ensure there is a more consistent and rigorous approach to tackling both the production and distribution of child abuse imagery. NSPCC research suggests that UK demand for child abuse material could be in the hundreds of thousands,[40] and the National Crime Agency has estimated 300,000 UK adults could pose a threat to children.[41]

Platforms must demonstrate the consistency and sufficiency of their response to child sexual abuse imagery (CSAI), including:

- the scope and effectiveness of their takedown processes;
- measures to proactively detect and disrupt new images being produced;
- a more proactive approach to removing images that might not meet the criminal threshold, but which have significant potential to cause harm.

Takedown processes for known images

Industry has developed well established takedown process, with mechanisms to identify and remove established illegal images, and to meet mandatory online child abuse reporting requirements. In the UK, the Internet Watch Foundation does excellent work.[42]

Although such takedown processes broadly work well, there are concerns about the consistency of the broader response. In 2018/19, the National Center for Missing and Exploited Children received 90 per cent of its reports from social networks from just one firm, Facebook.[43] This

38 This could also include the use of privacy-preserving approaches, including the use of on-device AI to detect concerning behaviour on children's devices.
39 TikTok took the welcome decision to restrict direct messaging to users aged 16 or under, although generally platforms have been reluctant to restrict or modify higher-risk design features.
40 Based on German research which estimates that 2.4 per cent of German males had seen child abuse material Jutte, S (2016) Online child sexual abuse images: doing more to tackle demand and supply. London: NSPCC.
41 National Crime Agency press release, 3rd March 2020.
42 Further information in the Internet Watch Foundation's annual reports.
43 Data from NCMEC.

suggests that other platforms are failing to report abuse material on anything comparable to the likely scale of the problem on their sites.

Perhaps most worryingly, some platforms do not appear to be enforcing takedown processes adequately.

The Canadian Centre for Child Protection, whose Project Arachnid tool has identified 6.1 million images since 2016, has found that some sites routinely refuse to comply with takedown requests of children aged as young as 9 or 10. Some platforms argue that if there is any (even very early signs) of sexual maturation, it is not appropriate for them to take down images, unless the age and identity of the child is already known.[44]

The Online Harms Bill should require platforms to have clear and consistent mechanisms in place to identify and remove known abuse images. The regulator must closely supervise the effectiveness of these processes. As part of its risk-based approach, it might usefully signal this will be a priority area for thematic review, and that it is prepared to take swift enforcement action against platforms that fail to deliver consistent and effective processes.

Platforms should be required to demonstrate consistent approaches to detect both still image and video-based abuse material. At present, many platforms arguably adopt a less comprehensive approach to scanning live and recorded video than they do for still images.[45]

It is vital that platforms are required to take a more child-centred, risk-based approach to their takedown processes. At present, while all social networks have clear policies that they do not allow abuse imagery, there is concerning inconsistency about how such policies are applied. The Canadian Centre reports this assessment is often 'highly subjective, inconsistent, and is cautious to the point of absurdity.'[46]

In order to discharge the Duty of Care, platforms should have clear processes in place to assess whether an image subject is likely to be a child, and should take all reasonable measures available to them to inform an age assessment. This includes the use of age-assurance technology, which will increasingly be used across a wide range of commercial and compliance functions.

Sites should be expected to take down abuse images where a hotline has judged the image is of a child, or on the balance of probability, the image is determined to be illegal material.

Tackling the production of new images

Online platforms account for a growing number of new images being produced, with abusers using social networks and gaming sites to coerce children into producing self-generated images, or to perform sexual acts on livestream sites.

According to the Internet Watch Foundation, self-generated imagery accounts for nearly one third of child abuse images, with three-quarters of such images featuring children aged 11-13. Although industry will claim progress has been made in the removal of child abuse images, not nearly enough has been to be done to tackle the production of child abuse at source.

Under any regulatory scheme, sites should therefore be required to invest in technology that enables it to proactively identify new abuse material, with processes in place to ensure such tools are applied consistently. Platforms must ensure such technology is rolled out across still images, video and livestreams.

Platforms should also have a broader responsibility to tackle the production of new images on their sites.[47] This responsibility could be discharged in a number of ways, for example through the adoption of intelligent design features, privacy preserving on-device measures, and easily accessible mechanisms to request the removal of self-generated content.

Although children have clear rights under GDPR to request the removal of any content, purely on the grounds of withdrawing consent,[48] platforms do not have easily navigable or child-centred processes that readily enable this. Children typically face an onerous and highly challenging user journey to request the removal of self-generated content that may have been posted or shared online.[49] Some children may feel disempowered to report self-generated content that is being shared or re-posted online, because they feel platforms may not take their report seriously enough. This complicated and

44 Canadian Centre for Child Protection (2019) How we are failing children: changing the paradigm. Winnipeg: CCCP.
45 As explored extensively by the New York Times in autumn 2019.
46 By their estimates, some large platforms have refused to remove CSAI associated with children they estimate to be 10 years of age. Canadian Centre for Child Protection (2019) How we are failing children: Changing the paradigm. Winnipeg: CCCP.
47 For example, through intelligent design features, investment in proactive technology, and through the adoption of a thorough risk assessment process.
48 Article 17 GDPR.
49 This despite the Information Commissioner's Office providing clear guidance that, in accordance with article 7(3), it is disproportionate to request identity documents when requesting takedown, if these were not required at the point of account creation. ICO (2018) Children and the GDPR: Guidance. Wilmslow: ICO.

often challenging process may significantly exacerbate the distress felt by young people, and may compound the risk that these images are shared widely, resulting in ongoing abuse, coercion and control.[50]

Child-centred, proactive approach to takedown processes

In December 2019, the Canadian Centre for Child Protection proposed a new framework[51] for tackling child abuse imagery, based around core principles that companies should be required to adopt a more child-centred and proactive approach to removing abuse material.

There is a compelling argument that the Duty of Care should necessitate a more proactive framework for the takedown of tackling child abuse images, but this must be delivered in a risk-based, and highly proportionate way.

Crucially, this approach should require platforms to identify and takedown images that may not meet the current legal threshold to be considered child abuse material, but which still warrant action. This is because they may facilitate access to illegal images; be used in a context for sexual gratification; or where failure to takedown images may perpetuate the impact on the children being abused.

Earlier this year, platforms agreed to take action on such images as part of the voluntary principles agreed with the Five Eyes Governments,[52] but this focussed on enhanced reporting processes, and stopped short of a formal takedown procedure.

Despite the clear abuse risks associated with such content, many firms 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, platforms focus their approaches in terms of illegal child abuse material that is seen by them to 'clearly and objectively meet a concrete definition'.[53]

It is vital that the regulator adopts a more child-centred approach that recognises that the systemic approach and processes used by platforms must not only result in the removal of clearly illegal content, but also action on the risks associated with material that could signpost to, or facilitate access of, illegal material.

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

Abuse image series may often appear, to anyone other than an offender, relatively innocuous but may quickly progress to a child being sexually abused. There appears to be growing demand among some abusers to collect the full series of abuse images.[54] In some cases, these are deliberately used by abusers because they anticipate such images won't be proactively removed by the host site.

Crucially, abuse sequences may also be used to signpost to or advertise illegal material hosted elsewhere, including on the dark web or on encrypted sites.

Such images effectively act as 'digital breadcrumbs' for abusers to locate other clearly illegal abuse material, and allow offenders to identify and form networks with each other.

Re-victimisation and image misappropriation

Platforms should be expected to take appropriate action to address material posted for innocent purposes, but which is misappropriated for the purposes of sexual abuse or re-victimisation.

In accordance with a risk-based approach, platforms should remove such images, where notified by a credible hotline that the image has been used for or to facilitate sexual abuse; and proactively scan for inappropriate or abusive conversations that may be related to pictures of children.

50 The NSPCC and Internet Watch Foundation have developed the Report Remove tool, which can support a young person to report an image shared online, and to enable the young person to get the image removed.
51 Canadian Centre for Child Protection (2019) How we are failing children: Changing the paradigm. Winnipeg: CCCP.
52 Home Office (2020) Voluntary principles to counter child online sexual abuse and exploitation.
53 According to Evelyn Douek (ibid 7), who notes there is a consensus among industry that the 'desirability and definition of child sexual abuse material is quite properly well settled', and that continual re-evaluation of the child abuse threat is therefore not necessary. However, the definitional parameters are far from settled – for example the Council of Europe's Budapest Convention defines fabricated images as illegal, but the US legal parameters do not, an issue which is likely to become more pressing with technological developments such as deepfake technology and the growth of artificial reality environments.
54 Based on discussions with the Canadian Centre and law enforcement agencies.

Test three: Legal but harmful content

If regulation is to succeed, it must tackle clearly inappropriate and potentially harmful content. This includes material that promotes or glorifies self-harm and suicide; and child sexual abuse imagery that in and of itself might not meet the threshold for illegality, but which signposts towards or facilitates access to illegal images.

In its interim response to the white paper, the Government set out that regulation will establish differentiated expectations for illegal content, and content that is legal but has the potential to cause harm. The regulatory framework will require companies to explicitly state what legal but harmful content and behaviour they deem to be acceptable on their sites, and to subsequently enforce these terms and conditions consistently.[55]

According to the interim response, all companies in scope will also need to ensure a higher level of protection for children, and to take reasonable steps to protect them from inappropriate and harmful content. However, the white paper did not set out further information on the intended approach.

As set out above, 'abuse image sequences' form a significant part of the child abuse threat – and perpetuate the impact of abuse on victims. The Duty of Care is unlikely to deliver its full potential unless it can effectively tackle both illegal and otherwise legal drivers of abuse, as part of a holistic, risk-based approach.

This also problematises the clear, but if seen through a harm-based lens unhelpfully simplistic, distinction between legal and illegal content that often characterises this debate.

We agree that Duty of Care should not focus on the removal of specific pieces of legal content,[56] but should tackle the means through which children are exposed to legal but harmful content through design features, algorithmic recommendation and content amplification. Platforms must adopt a systematic approach to the enforcement of their terms and conditions.

However, the regulator must also adopt a child-centred, harm-based approach to the development of its regulatory scheme. The scheme must clearly recognise the risks to children, and balance these appropriately against freedom of expression.

This approach should reflect that freedom of expression is not absolute, and that under Article 8 of the European Convention on Human Rights, states have a positive obligation to secure the physical and psychological integrity of an individual from other persons.[57] This applies particularly to the well-being of vulnerable groups, and in order to protect their right to a private life, includes the protection of a child from physical and mental harm.[58]

In a scenario where images of a young person's abuse are allowed to remain in circulation on platforms and this causes acute psychological distress or harm to the child, questions would arise about compatibility with Article 3 (which includes the right not to be subjected to inhuman or degrading treatment). States are under a duty to take action to address known risks (including where the state is not the primary violator) and ensure adequate legal structures and sanctions are in place to protect children from sexual abuse and harm.[59] States need to take steps to fulfil their positive obligations to protect children from abuse which occurs on or offline.

Regulation must recognise that the potential for harm cannot be understood solely in terms of the legality of behaviour or material. In the case of the most egregious legal harms, including self-harm and suicide material, there is a clear precautionary basis on which to act to protect vulnerable children against the risks associated with the amplification or algorithmic recommendation of it.

55 HM Government (2020) Interim response to the Online Harms white paper.
56 Excluding material which directly supports child abuse.
57 European Court of Human Rights (2020) Guide to article 8: right to respect for private and family life, home and correspondence. Strasbourg: ECHR.
58 KU vs Finland. European Court of Human Rights (2015) Internet case law of the ECHR. Strasbourg: ECHR.
59 O'Keeffe v Ireland. European Court of Human Rights, Grand Chamber, Application Number 35810/09, 28 January 2014.

The most serious legal harms continue to affect children at scale, and underline why action to protect children so essential. Facebook's own figures suggest that up to 5 in every 10,000 views contain prohibited material that glorifies and promotes self-harm and suicide.[60] For vulnerable children being served up such content through algorithmic profiling, this is likely to be a significant underestimate.

The Government must therefore remain live to the risk that if it is left to platforms to set their own terms and conditions, this creates a moral hazard for them to set the bar as low as possible – and could reinforce the existing dynamics in which platforms 'hold all the cards'.

Any differential application of the Duty of Care would fail children if it poorly reflects, or is unable to adequately respond to, the full extent of online harms to which they are likely to be exposed. A 'two track' approach cannot result in children diminished protection, even if this is not the intention.

As part of its risk-based approach, the regulator should signal it intends to subject platforms to enhanced regulatory scrutiny, in the form of thematic reviews. It should also provide a non-exhaustive list of examples that provide broad guidance to industry on its expectations for a systemic approach to tackling legal but harmful risks to children.

60 Facebook transparency reports, available on Facebook's website.

Test four: Transparency, investigatory and disclosure powers

For too long, social media companies have been able to selectively disclose what, if anything, they do to protect children from abuse on their platforms.

There is no requirement on tech firms to disclose the scale and extent of abuse risks on their sites. For example, during the pandemic Facebook seemed either unable or unwilling to answer the Digital, Culture, Media and Sport Select Committee's questions on the impact of lockdown on their moderation efforts.[61]

Larger platforms issue their own transparency reports in the form and frequency they consider appropriate, but crucially 'no external agency has the ability to assess objectively whether platform approaches to address online harms are effective'[62]. At present, no platform reports on the risks faced by UK children specifically.

Transparency is crucial to the regulator's work – and is arguably as important as enforcement powers that inevitably tend to attract more attention. As the Tony Blair Institute for Global Change puts it: 'Enforcement mechanisms may not be as good a long-term solution to the safety of people online as a close relationship between regulator and platform, which includes transparency and scrutiny on the regulator's terms.'[63]

At stake is the existing 'extreme information asymmetry' between platforms and regulators,[64] which leaves regulators with practically them same level of information as a user. Unless this asymmetry is overcome, there is an inevitable risk that the regulator will lack the necessary information to carry out objective analysis, and to identify systemic harm.

If the regulator lacks such power, there is a clear potential it is required to take decisions on low quality evidence, which in turn could be subject to legal challenge; and that it becomes highly risk-averse and incapable of taking effective regulatory action.[65]

The Government therefore needs to be suitably ambitious in the powers it gives the regulator – if it doesn't have the necessary powers to 'lift the lid' on company performance, it risks becoming a paper tiger.

In order to address these risks, the Government must:

- Give the regulator wide-ranging and comprehensive powers to require information disclosure. Platforms should be made to disclose any information that the regulator considers necessary, either to assess its investigations or its ongoing work;
- Provide the regulator with clear and unequivocal powers to launch investigations and investigate evidence of non-compliance. The Government must ensure the regulator has the appropriate resources to deploy its powers, and to tackle highly technical and complex drivers of harm;
- Balance regulatory powers with new disclosure duties on firms. These duties should require platforms to risk assess their products and services, proactively share information on new and emerging safety risks, and incentivise them to notify the regulator of any aspect of its service of which it could expect to be made aware – this should include safety breaches.

Transparency reports

Transparency reports can be an important part of the regulatory solution, allowing the regulator, civil society and users to fully understand industry processes and hold them to account. This could also support a 'race to the top'.

However, such reports must provide meaningful and interrogable data, and demand metrics that are measurable, comparable and that incentivise improvements to platform processes and design. In order to provide such meaningful data, regulatory reporting should emphasise the impact of actions that

61 Digital, Culture, Media and Sport Select Committee (2020). Second Report of Session- Misinformation in the COVID-19 Infodemic. London House of Commons.
62 Douek, E. (2020) The rise of content cartels: Urging transparency and accountability in industry-wide content removal decisions. New York City: Knight First Amendment Institute, Columbia University.
63 Beverton, Palmer, M (2020) Online harms: Our View on the UK Government Plans. London: Tony Blair Institute for Global Change.
64 Loutrel, B (2019) Creating a French framework to make social media platforms more accountable: Acting in France with a European vision. Paris: Direction interministérielle du numérique et du système d'information.
65 Beverton-Palmer, M (2020) Cited above.

minimise or eliminate risk,[66] rather than just the number of actions taken.[67]

It is appropriate that the Government engages widely on the development of its reporting framework, including through its Transparency Working Group. However, it must ultimately be for the regulator to determine the eventual composition of reporting requirements – with decisions on disclosures ultimately being driven by the public interest, not that of the companies.[68]

There are understandable trade-offs involved in the level of transparency that should be expected, for example the risk that data enables bad actors to game the system. However, given 'the status quo is very far from the optimal balance', and arguably amounts to a form of 'transparency theatre',[69] it will be important that the scope and demarcation of reporting requirements actively builds, and doesn't weaken, trust in the regulator.

It is appropriate that the regulator adopts a tiered approach to transparency reporting, with larger platforms being expected to report on a more comprehensive basis than smaller firms.

We support proposals for transparency reports to be externally audited.[70] This will build confidence in the quality and robustness of regulatory disclosures, and minimise the risk that platforms seek to present data in a selective or potentially inaccurate way.[71]

Investigatory and information disclosure powers

It will be essential for the regulator to be equipped with robust and intelligently designed investigatory and information disclosure powers.

Ofcom has broad existing information disclosure powers under s135-146 of the Communications Act 2003. These powers allow the regulator to request any information necessary for it to carry out its functions, but must be proportionate to the use to which the information is to be used.

These comprehensive powers allow Ofcom to make well-evidenced decisions that secure broad confidence in its actions, including businesses in regulatory scope.

The regulator should have clear powers to investigate platforms for non-compliance. These must include the power to require documents and other information, to carry out searches, and to instruct platforms to carry out impact studies (as is required in the tobacco sector).

Ofcom should be able to commission a 'skilled person' review, where it has concerns about a platform, or it requires further understanding of the adequacy of its systems and processes.[72] This is a highly effective measure that is used to support investigations in financial services regulation, and addresses the challenge it faces to recruit a sufficient volume of staff with the requisite technical skills required to effectively discharge its supervisory and enforcement functions.

Drawing on the powers available to the Financial Conduct Authority,[73] the regulator should be able to commission such a review, and where necessary to directly appoint the 'skilled person' to conduct it, with the regulated party being liable for the costs incurred.

Proactive duty on platforms to disclose

Platforms should be subject to a general proactive duty to disclose information to the regulator that it could reasonably be expected to be informed about. This will act as an important means of regulatory intelligence-gathering – and perhaps more importantly, is likely to be a useful means of hardwiring regulatory compliance into sites.

Although potentially broad, the scope of this duty can be drawn with sufficient clarity that social media firms can properly understand their requirements. This will ensure the regulator is not inundated with (and platforms are not bombarded by) unmanageable and unhelpful volumes of reporting.

A similar proactive duty already applies in the financial services sector. Principle 11 of the financial services regime requires firms to deal cooperatively with the regulator, and to disclose anything of which the regulator would reasonably expect notice. This is supported by a non-exhaustive list of examples.

66 Current transparency reporting tends to emphasise the publication of metrics, but without contextualised information that enables an assessment of the resulting impact and scale of platform response. See Evelyn Douek's analysis of hashing metrics in the GIFCT Transparency reports. Douek, E. (2020) The rise of content cartels: Urging transparency and accountability in industry-wide content removal decisions. New York City: Knight First Amendment Institute, Columbia University.

67 This approach will also support freedom of expression, through reducing the perverse incentive for platforms to remove excessive pieces of content.

68 ibid.

69 ibid.

70 Beverton-Palmer, M et al (2020) Online harms: bring in the auditors. London: Tony Blair Institute for Global Change.

71 Platforms have previously been found to have underreported regulatory filings, for example Facebook was fined 2 million euro in July 2019 for under-reporting complaints on hate speech in relation to the Netz DG regulations in Germany.

72 See the Financial Conduct Authority's website for more information: https://www.fca.org.uk/about/supervision/skilled-persons-reviews.

73 Under section 166 and 166a of the Financial Services and Markets Act.

'Red flag' reporting where children's safety is compromised

At present, there is no requirement for platforms to report in the event of significant lapses in systems and processes that compromise children's safety or could result in them being at risk. However, such reporting is widely used in other regulatory regimes.[74]

Platforms should no longer be able to self-police in this way. We therefore propose that platforms be subject to 'red flag' reporting. This requires immediate disclosure to the regulator in cases where the safety or well-being of children could be compromised, or where there has been a significant material breach in child safety processes.

Duty to risk assess new products and services

Platforms should be required to conduct risk assessments before launching new products and services, or making significant changes to existing ones, and to share these with the regulator prior to services being launched in the UK.

Impact assessments should specifically consider the potential impacts of services on children, and enable a platform to demonstrate to the regulator that it has taken all appropriate measures to assess and mitigate all reasonably foreseeable risks to children.

This measure will incentivise companies to embed the Duty of Care across the product lifecycle, and to critically assess the likely impacts of a product before it launches – the antithesis of the 'move fast and break things' approach which has led to child-facing risks often being treated as a secondary concern.

Compliance with information requests

Platforms should be incentivised to comply with information requests, and accordingly the regulator should credit timely disclosure of information that may lead to subsequent enforcement action.

Ofcom should also have the power to impose sanctions for non-compliance with information requests or attempts to provide misleading data.[75] Although the financial impact of sanctions might be limited, in the case of large platforms this could deliver reputational effects.[76]

74 For example, financial services companies are required to make reporting disclosures under the anti-money laundering and financial services regime, and licensed gambling firms must report breaches against self-exclusion protocols.
75 Ofcom has the existing power to issue fines for non-compliance with information requests under the Broadcasting Code.
76 Centre for Data Ethics and Innovation (2020) Online targeting: final report and recommendations. London: HM Government.

Test five: Enforcement powers

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

This reflects the principle that the platforms that create risks should be responsible for the costs of addressing them. For too long, children, families and society have been left to bear the costs, in the devastating form of the emotional, mental, and physical, but also social and economic costs, of child abuse.

The regulator must be given a comprehensive package of compliance and enforcement powers that incentivise behavioural change in companies that might otherwise continue to put children at risk. Any sanctions regime must be proportionate to the size and scale of the companies in scope. Given the scale of the largest platforms, this means the magnitude of sanctions must be significant.

The Government must deliver the enforcement measures that are required. Unless we see a comprehensive package of measures that provide a strong deterrence effect, and can adequately focus the minds of senior managers, the Duty of Care regime might fail.[77]

Building an effective enforcement regime

Given the serious nature of the harms in scope, the weak economic incentives for compliance, and the global size and structure of the biggest online services, the regulator will require a robust set of enforcement mechanisms.

We envisage a range of enforcement options, both civil and criminal, should be made available to the regulator. Crucially, these should apply both to the corporate entity, but also senior managers with responsibility for ensuring a platform's Duty of Care responsibilities are met.

It is likely the regulator will actively need to draw on a broad range of levers to incentivise and necessary enforce compliance. If the regulation is to succeed, each of the following measures must therefore be built into the regulatory regime.

Financial sanctions

The regulator should be able to impose financial penalties where there is a breach of the platform's Duty of Care, or in circumstances where a platform fails to cooperate with the regulator or is considered to have provided misleading information to it.

Financial penalties must be of sufficient magnitude to deter non-compliance, and to eliminate any financial gain or benefit from a platform's decision not to comply with its regulatory requirements.

For the most significant breaches, for example a platform that consistently fails to deliver against its Duty of Care, sanctions should be levied on a similar magnitude to GDPR, i.e. up to 20 million euro, or 4 per cent of global turnover.

However, it is important the limits of financial penalties are clearly understood. Even maximum financial penalties may have a limited impact on the major platforms, given their global revenue. The largest technology companies have billions sitting in the bank as cash at hand, and making

no return for shareholders. In this context, the micro economic effect of fines is blunted as they will have little impact on the marginal behaviours of either the management team or shareholders.

In any event, 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.[78]

Senior Managers Regime

There is a clear benefit in ensuring that responsibility for regulatory compliance is held at the most senior levels of social media companies. It is therefore essential the Government legislates for a Senior Managers Regime.

Across other regulated sectors, senior management liability is widely seen as a valuable means of securing regulatory compliance, securing a solid risk and control culture in regulated firms, and as a powerful means of delivering both organisational and sectoral cultural change.[79]

77 In oral evidence to the Digital, Culture, Media and Sport Select Committee, the Culture Secretary Oliver Dowden described some of the measures originally proposed in the Online Harms White Paper, which are essential to the regulator's success, as 'draconian.'

78 Centre for Data Ethics and Innovation (2020) Online targeting: final report and recommendations. London: HM Government.

79 Chiu, I (2016) Regulatory Duties for Directors in the Financial Services Sector and Directors' Duties in Company Law - Bifurcation and Interfaces. *Journal of Business Law, 2016.*

Under the Senior Managers Regime, personal liability would apply for senior management with a 'significant influence function'.[80] Senior Managers would be subject to a set of conduct rules, which would reinforce corporate-level requirements on platforms and incentivise senior management decision making to internalise the Duty of Care in the delivery of their functions.

Conduct rules could reasonably include a requirement to:

- Take reasonable steps to ensure the business is controlled effectively;
- Take reasonable steps to ensure all business functions for which a manager is responsible comply with relevant regulatory requirements;
- Ensure any delegation of responsibilities is to an appropriate person, and that the discharge of these functions is overseen correctly;
- disclose appropriately any information of which the regulator would reasonably expect notice.

If the senior manager failed to identify reasonably foreseeable risks, and ensure their platforms had appropriate polices and protections to mitigate them, this could reasonably be considered as a conduct failure that knowingly contributed to a regulatory breach.

Under such circumstances, compliance action could result. This could take the form of the Director being fined, or disbarred from taking on similar regulated roles.

For the most serious of breaches, the senior director could be found criminally liable for the consequences of failing to discharge their responsibilities. The named director could also be considered to have committed an offence under the Company Directors Disqualification Act.[81]

Corporate criminal responsibility

Criminal sanctions should apply in the event that a social media provider commits a gross breach of the Duty of Care. Such offences would be proportionate, reasonable and clearly linked to regulatory objectives.

The principle that a corporate entity should face criminal sanctions is well-established in law, and it is both logical and necessary to extend this precedent to tackling online harms. For example, the offence of Corporate Manslaughter has been part of UK law since 2008, where a prosecution may be brought if failings by an organisation's senior management are a substantial element in any breach of the duty of care that it owes to employees or the public, and this results in death.

Other regulated sectors already make provision for corporate criminal sanctions to apply in the event there are significant, systemic failures that that result in serious harm or criminality. Criminal charges can be brought where there are repeated and persistent breaches under the Health and Safety Act 1974. Strict 'failure to prevent' offences exist in relation to bribery and tax evasion.[82]

Corporate criminal sanctions are likely to act as a strong deterrence, and in the event that an offence took place, would deliver strong adverse reputational effects. Drawing on existing legal precedent, a corporate criminal offence could occur where a platform grossly failed to discharge its Duty of Care to address harms facilitated or enabled by its service.

In such cases, if a court found that the platform had failed to introduce procedures or that these were not discharged adequately, it could determine that this constituted a gross breach - and could result in a corporate conviction.

We envisage that charges would only occur in extreme situations, but that the extension of corporate criminal sanctions into the online harms regime will help to embed regulatory compliance at the highest levels. As in other sectors, it will frame businesses' approach to managing risk, and any prosecutions would publicly underline the severity of failing to discharge the Duty of Care towards children.

Enforcement notices and reputational remedies

The regulator should be able to direct sites to apply remedial measures in respect of children's safety, for example requiring the adoption of specified safety-by-design features. Where proportionate, it should also be able to prohibit the continuation of certain activities, for example restricting the use of certain features.

Reputational remedies, for example the use of public censure and adverse publicity orders, could prove effective. This could include press notices to raise awareness of enforcement action, or instructions being given to a platform to display a prominent message on its home screen setting out the details of how a regulatory breach put its users at risk.

80 Drawing on the model adopted in financial services regulation.

81 The Company Directors Disqualification Act is increasingly being used to ensure named corporate responsibility for legal and regulatory obligations. For example, last year the Government consulted on proposals to bring forward an offence under the Act, where a business fails to prepare a Modern Slavery Statement in accordance with its obligations under the Modern Slavery Act.

82 The Criminal Finance Act 2017 created new corporate criminal offences where a company can be prosecuted if it is unable to show it has sufficient controls in place to prevent staff from committing an initial offence.

According to PA Consulting, publicising reputational breaches and enforcement action is an important means of building regulatory awareness and trust – and so could potentially support parents in being better informed about the risks posed by sites. This research shows that consumers feel more protected when they've heard of the regulator (82 per cent) and when regulatory breaches are publicised (80 per cent).[83]

Disruption to business services

It may be necessary to consider issuing a notice to Internet Service Providers (ISPs) to block access to services that persistently fail to engage with the regulator, or that pose a significant safety risk.

While ISP blocking is likely to be used in extremis, it may be necessary to grant the regulator this power, particularly for smaller, extraterritorial sites.

Risk-based approach to enforcement

We envisage the regulator would adopt a risk-based and proportionate approach to enforcement, and would seek to apply its sanctions powers judiciously.

This should address concerns that the sanctions regime may be too drastic; could inadvertently entrench the competitive position of the current market giants; and could deter service providers from offering social media services, in turn having a chilling a chilling effect on freedom of expression.[84]

While it is likely the regulator would take the strongest enforcement action against platforms that failed to tackle the most egregious illegal content, the regulator should not adopt a bifurcated approach to enforcement powers[85] – which would prevent the regulator from being able to take stronger forms of enforcement action against breaches relating to legal but harmful content.

Any differentiated approach could disincentive platforms from tackling issues relating to legal but harmful content, particularly if it resulted in a constrained set of sanctions and/or enforcement appetite, and it would not be consistent with a child-centred approach.

83 PA Consulting (2018) Re-thinking regulators: from watchdogs of industry to champions of the public. London: PA Consulting.
84 These arguments are explored further in Damian Tambini's assessment of the Online Harms White Paper. Tambini, D (2019) Reducing Online Harms through a Differentiated Duty of Care: A Response to the Online Harms White Paper.
85 As proposed in Tambini's paper.

Test six: User advocacy arrangements

The Online Harms regulator will protect children most effectively if it is able to develop deep relationships with civil society, industry and other regulators. However, how civil society can engage with the regulator, and the mechanisms and support needed for this, is a crucial if largely overlooked issue.

Up to now, the Government's emphasis has understandably been largely on the design and function of the regulator. However, lessons from other regulated sectors demonstrate the importance not only of regulatory design and powers, but of securing a wider regulatory settlement – a landscape in which user advocacy arrangements are appropriately empowered to act on behalf of service users.

If Online Harms regulation is to succeed, there is a compelling case for the Government to adopt formal statutory user advocacy arrangements, funded by the industry levy. This would enable strong and coherent representation on behalf of children, and provide the necessary counterbalance to well-resourced industry engagement.

Why do we need user advocacy arrangements?

There is a powerful argument for user advocacy arrangements that promote and protect the interests of children – and that ensure children's needs are surfaced to the regulator. In any effective regulatory regime, there will be a need for user advocacy arrangements that can:

- Represent the interests of children, including in regulatory debates;
- Undertake research to assess and identify risks faced by children online, meeting the evidential standards required by a regulator, and to;
- Highlight areas where the regulator needs to act, using its expert understanding of how online harms affect children, and children may face differential or enhanced risks using online services.

User advocacy arrangements would broadly mirror the regulatory settlement developed in other sectors, with well-established user advocacy arrangements in place for the utilities and essential services sectors, and most recently, a commitment to legislate for a user advocate on telecoms.[86]

The industry levy is an appropriate mechanism for funding user advocacy arrangements – this is entirely consistent with the well-established 'polluter pays' principle, and it is a wholly proportionate and reasonable set of costs when set in terms of the commercial return available to platforms that offer their services to children, but do not protect them adequately.

A user advocate for children

Even if the regulator is equipped with the legal powers it needs, and is able to secure the necessary technical and sectoral expertise to discharge its functions, a strong, credible and authoritative voice to represent children's interests is vital.

At present, a range of civil society organisations represent children. However, it cannot be taken for granted that civil society can continue to perform these activities either in perpetuity, or to the level and extent that is necessary to support, and where necessary to offer challenge, to the regulator.

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

A user advocate would be able to effectively navigate and articulate the particular safeguarding challenges in respect of online harms and young people; meet the need to identify child abuse issues in a rapidly changing market, and could serve as an effective counterbalance to outsized industry resourcing and influence.

Children face rapidly emerging forms of harm

Ofcom will be inheriting regulatory responsibilities in a sector which is characterised by rapid technological and market change, and a correspondingly agile and complex child abuse threat.[87]

86 Department for Digital, Culture, Media and Sport (2020) A New Champion for Mobile and Broadband Customers.
87 WeProtect (2019) Global Threat Assessment, prepared by PA Consulting.

Innovation brings new opportunities for children, including the growth of livestreaming and video chat products, but repeatedly we see the safeguarding implications are often poorly understood.

In the context of such a rapidly evolving threat, it will be vital that the regulator is able to be informed by credible, authoritative and well-resourced civil society expertise – and that safeguarding concerns can quickly be identified, for example through dedicated safeguarding expertise and networks, which the regulator itself is unlikely to possess.

The pace of change also means that by the time the regulator may have been through the lengthy process of identifying evidence of harm, consults on remedies, and decides to take action,[88] children may have already been exposed to considerable risk of abuse. Effective user advocacy arrangements would provide an expert early warning function, and so could reasonably mitigate this.

Ensuring children's interests are fairly balanced against industry

Tech firms are a well-resourced and powerful voice, and will seek to exert strong influence when decisions are made about their services. It is highly likely that industry may seek to prevent the regulator from building a full understanding of the impact of their services on children.

Larger platforms are highly likely to consider protracted and expensive legal action to frustrate or challenge regulatory decisions. This could reasonably be expected to influence the regulator's risk appetite, which could result either in delayed action or a reluctance to proceed with more ambitious ex ante initiatives.[89]

Powerful industry interests are not unique to the tech sector, but the size of and resources available to the largest players is arguably distinct. In the development of online harms regulation, there is a balancing act between allowing the proposed regulatory duties to promote innovation,[90] and ensuring children – perhaps the most vulnerable of all user groups - are protected.

In most other regulated markets, these risks are addressed through strong, independent advocacy models that can provide appropriate counterbalance. Without such arrangements in place for online harms, there is a clear risk that children's interests will become asymmetrical, and unable to compete effectively with the arguments made by industry. In turn, this could jeopardise the regulator's ability to create better outcomes for children.

The regulatory parameters of online harms will not include some functions performed by other advocacy bodies, but there is a clear impetus to develop advocacy and representation arrangements that create a 'level playing field' for children.

Children face distinct and enhanced risks – but user redress is a poor answer

The Online Harms White Paper is predicated on the basis that children face pronounced risks both because of both their inherent vulnerability; and because it is reasonable to expect they may be less aware of the safeguards concerned, and their rights in relation, to online platforms.

In respect of online harms faced by adults, the Government is minded to propose a three-fold blended approach that delivers regulatory action; promotes the growth of safety-by-design solutions; and promotes user empowerment.[91]

However, the nature of the risks faced by children, and the inherent difficulty in developing user empowerment initiatives that could reasonably be expected to benefit them,[92] means that Government will be unable to draw on this blended approach for children.

For example, there are clear obstacles to being able to deliver effective redress options for children, and in turn, to use the outcomes of redress mechanisms to assess platform risks and identify issues that would benefit from regulatory attention.

The nature of many online harms may not be readily recognisable to either children or adults, for example if a child is being served harmful content as a result of algorithmic profiling or through tailored design choices.

Many children who are experiencing online abuse, for example children that may be groomed on social networks, may not readily recognise their experience as such.

It therefore becomes important to have strong advocacy and representation structures that ensure children's issues are appropriately represented, and to ensure child-facing risks can be appropriately surfaced to the regulator.

88 The issue of extended regulatory timescales is set out well by Citizens Advice in their assessment of sectoral regulators. Citizens Advice (2018) Access denied: the case for stronger protections to protect telecoms users. London: Citizens Advice.
89 ibid.
90 In the Online Harms White Paper, the Government sets out plans for a statutory duty on Ofcom to give regard to innovation.
91 Comments made in the Westminster eForum Online Harms conference, June 2020.
92 There is a significant 'cognitive burden' associated with user empowerment on online services, which will likely be heightened for children and young people. Centre for Data Ethics and Innovation (2020) Online targeting: final report and recommendations. London: HM Government.

Supercomplaints

There is clear merit in the adoption of supercomplaint powers, which would allow designated bodies to raise complaints about systemic issues relating to online harms.

Our preferred approach is to grant designated supercomplainant bodies the right to raise a complaint about any feature, or combination of features, of a product or service that has the potential to cause harm to users. This would broadly reflect the powers available to designated bodies in the utility, telecommunications and consumer sectors, which can bring supercomplaints under section 11(1) of the Enterprise Act 2002.

Under section 11(6) of this Act, supercomplaints can only be tabled by bodies who appear to the Secretary of State to represent the interests of users (a similar designation framework applies for policing supercomplaints). We envisage similar designation should apply here.

Other considerations

In line with other regulatory examples, the regulator should have a duty to consult with expert groups in the exercise of its functions, including user advocacy arrangements, law enforcement and civil society groups.

The regulator should have a specific duty to assess the risk of harms to particular groups of users, and to assess how online harms may be disproportionately experienced by them.[93] This should include a consideration of how online harms may be differentially felt by users with one or more protected characteristics under the Equality Act.

Provision should be made for the regulator to be informed by a wide plurality of user experience. We recommend that the regulator develop user representation structures, enabling it to inform its approach through engagement with those who have experienced online harms, and that represent a broad cross-section of UK users (including those that may be exposed to risk on an intersectional basis).

93 For example, during lockdown there was evidence that people with long-term health conditions were being targeted online e.g. users with epilepsy were targeted with content designed to trigger seizures. Similarly, there is extensive research which suggests that LGBTQ+ children are likely to face greater levels of harassment and abuse online, and are more likely to be contacted by people online who aren't who they claim to be. Research for Brook and the National Crime Agency has shown that LGBTQ children may be exposed to additional risks online. McGeeney, E; Hanson, E (2017) Digital Romance: a research project exploring young people's use of technology in their romantic relationships and love lives. London: Brook.

Regulatory expectations on high-risk and emerging design features

Platforms should face a range of wider requirements to effectively tackle the child abuse threat, and the regulator should incentivise the adoption of 'safety-by-design' as a core principle for online services.

If platforms introduce features designated as high risk, including livestreaming, private messaging and end-to-end encryption, they must demonstrate that appropriate safeguards are in place. If a site introduces high risk functionality, but cannot demonstrate reasonable mitigations are in place, this would not be consistent with the Duty of Care.

Minimum safeguarding standards

Platforms should be required to adopt a minimum set of safeguarding standards, in the form of a 'safety-by-design' Code of Practice.

In the same way that food safety legislation must apply equally to the local sandwich shop and the largest supermarket chains, it is right that children receive a consistent set of minimum protections, regardless of the social networks they use.

Minimum protections should include: default privacy and safety settings for children's accounts; accessible, age-appropriate explanations of terms and conditions; a transparent and responsive complaints process; and a dedicated reporting flow for complaints that relate to child abuse.

Platforms should be required to take reasonable steps to determine the age of their users, through the use of age-assurance processes, to enable additional safeguards to be applied to children's accounts.

The NSPCC favours the use of less intrusive age-assurance mechanisms, rather than the introduction of explicit age verification checks, for social networks and gaming services. Age assurance processes encompass a form of initial and ongoing 'know your user' checks, and should enable platforms to identify children so their accounts can receive the protections outlined above.

This represents a more targeted and proportionate response than age verification checks, which could be fraught with technical and data privacy challenges, and draw resources from safety risks that equally affect children aged 13 and over. Age verification is arguably a punitive response that penalises children and young people for the shortcomings of social media platforms - and shifts the onus from a reasonable expectation that sites identify and mitigate harms, to preventing children from being able to access online services they want to use.

High-risk design features

Online providers should be subject to additional regulatory requirements, and reasonably expect higher levels of regulatory scrutiny, if they offer higher risk design features.

The regulator should maintain a list of high-risk design features and update this regularly. Platforms should be expected to risk assess how high-risk features operate on their services, and demonstrate that the functionality is safe for children to use. If a platform cannot demonstrate that appropriate risk mitigations are in place, it should consider whether it is appropriate to continue offering it.

Companies will decide how they will mitigate the risks of high risk functionality, but in accordance with a risk-based approach, the regulator should actively assess the efficacy of these approaches, and conduct regular thematic reviews.

Livestreaming and video-chat services present high risks to children, including the risk of online grooming and the production of new child abuse material.

Our research underlines how the live, visual and inherently unpredictable nature of livestreaming services puts children at particular risk of online grooming – 1 in 20 children have been asked to remove clothing when livestreaming, rising to one in then children using video chat sites.[94]

Despite the rapid ongoing investment in content moderation technology,[95] it is arguable whether many sites have yet developed a suitably comprehensive approach to tackling child abuse on livestreams and video chats. Duty of Care regulation should incentivise companies to ensure a more coherent response.

94 NSPCC (2018) Livestreaming and video chatting – a snapshot.

95 Following the Christchurch attack, we saw platforms develop a better cross-platform response to taking down viral content, for example following the Halle synagogue attack. More recently, platforms have been tested with the September 2020 livestreamed suicide of a man on Facebook Live. Apparent shortcomings in Facebook's moderation of the livestream meant platforms including TikTok faced a cat and mouse exercise to remove the content. Gilbert, D (2020) Facebook Refused to Take Down a Live-Streamed Suicide. Now It's All Over TikTok. Published by Vice News.

Although further investment in technology is required, platforms could reasonably discharge their Duty of Care through the adoption of potential mitigations that include:

- metadata analysis to identify suspicious patterns of user behaviour, for example unusual spikes in views of livestreams by children and young people;
- real-time moderation of video streams, using nudity detection and age assurance classifiers to detect and disrupt online grooming;
- adopting a risk-based approach to design, for example preventing new users from livestreaming until a 'cooling off' period is completed, or allowing children to livestream only to their followers;
- avoiding high risk design features on video-chat sites, for example functionality which allows users to display their interests, or filter other users by shared interests;
- promoting livestream or video chat services to children only if suitable mitigations are already in place. In recent weeks, Facebook has been actively promoting its Messenger Rooms video chat service prominently on its homepage, despite the service carrying high risk design features.

Private messaging

The Duty of Care should extend to private messaging services, to tackle the clear risks associated with technology-affiliated grooming, and the sharing and distribution of child abuse images. It should be considered a high-risk design feature.

Any proposal to impose proactive scanning must be limited, clearly justified and subject to appropriate safeguards – however, it is entirely proportionate to scan material for child abuse imagery, and where reasonable risk thresholds are met,[96] to analyse an account's private messages for evidence of online grooming.

Although some platforms already proactively scan private messages, a consistent set of requirements is highly desirable. This will mitigate the risk that offenders are incentivised to migrate to online services that adopt less rigorous proactive scanning.

End-to-end encryption

End-to-end encryption will place children at palpably heightened risk of technology-facilitated abuse, and make it harder, if not impossible, for platforms to identify and disrupt child abuse material and grooming on their sites.[97]

Government should explicitly require the regulator to assess the impact of end-to-end encryption, and if a platform cannot demonstrate it has adequately mitigated the associated risks, it should be prevented from proceeding (or continuing) with it.

In its assessment of proposed mitigations, the regulator must explore the likely impact, efficacy and comprehensiveness of solutions.[98] Platforms must be able to demonstrate they can fully discharge their Duty of Care to children. They should not be able to balance partial mitigation of child abuse risks against wider societal benefits, for example improved user privacy.

Some platforms might attempt to 'game' the duty of care legislation, for example by making significant changes such as end-to-end encryption that could weaken children's safety prior to the regulation taking effect. Government should therefore adopt a clear position on end-to-end encryption in its interim Code on CSA, and when tabling the Online Harms Bill, make it explicitly clear the regulator has powers to retrospectively assess design choices that put children at risk.[99]

Cloud hosting providers

Large firms are failing to rollout consistent child abuse scanning processes across all their products, leading to divergent approaches across companies that offer social networks, messaging and cloud storage services.[100]

Despite the rapid growth of cloud-based hosting services, the majority of cloud providers fail to proactively scan for child abuse images at point of upload. This creates a clear, and unacceptable, distinction between the approach taken by tech companies when proactively scanning content on their email or social networking products, and the more limited approach for their cloud-hosting services.

96 For example, where there are reasonable grounds to assume an account is engaged in grooming activity as a result of metadata analysis or through behavioural or linguistic artificial intelligence.

97 Analysis from the National Center for Missing and Exploited Children suggests encryption could result in the loss of 70 per cent of Facebook's child abuse reports. In 2018, these reports resulted in 2,500 arrests and 3,000 children being safeguarded in the UK.

98 In July 2020, Facebook were unable to explain to the Home Affairs Select Committee during oral evidence how they would be able to detect child abuse once its services are encrypted, leading Chair Yvette Cooper to state: "I don't know if you recognise quite how serious it sounds that you have made a decision to go ahead with something and you don't seem to have any idea of how you are going to solve this massive problem about how to protect children."

99 Another example of risks that may pose first order risks to safety are proposals for social networks to move to decentralised models, which similarly to encryption, effectively engineer away the ability to perform content moderation. York and Zuckerman observe that decentralization as a cure for the concentration of power in the major platforms "replace[s] one set of moderation problems—the massive power of the platform owner—with another problem: the inability to remove offensive or illegal content from the internet.' York, J; Zuckerman, E. (2019) Moderating the Public Sphere, in Human rights in the age of platforms 137, 140 (Rikke Frank Jørgensen ed., 2019).

100 New York Times (2019) Child abusers run rampant as tech firms look the other way, published 9th November.

For commercial reasons, providers seem keen to reassure cloud users of privacy in their policies.[101] However, it seems difficult to square how a platform can decide it is legitimate to scan an image if it shared on their email service, but it is overly intrusive to scan the same piece of content when uploaded to the cloud.

Cloud storage providers should therefore be subject to the provisions of the Duty of Care, and should be expected to take reasonable measures to detect child abuse material and prevent hosting it.

Cross-platform collaboration

The regulator should set out clear requirements for industry to collaborate on child abuse risks. A duty to collaborate would enable tech firms to play their part in combatting the way in which harms spread rapidly through the social media ecosystem, and demonstrate an industry-wide commitment to tackling abuse.[102]

Platforms should reasonably be expected to develop mechanisms to share relevant expertise and intelligence on emerging threats; collaborate on the development of new technical solutions; and share data on constantly evolving abuse risks. Larger sites should reasonably be expected to contribute more towards cross-industry initiatives.

There are particular benefits to the sharing of industry datasets. Developing artificial intelligence tools for content moderation at scale is often hard and resource-intensive. In some cases, it may not be possible without access to large datasets to which only the biggest platforms currently have access.[103]

A largely siloed approach to tackling risks cannot adequately respond to the ways in which online harms proliferate. Child abusers can exploit multiple different platforms, although technology companies' responses remain mostly focused on their own sites.[104] A requirement for industry cooperation should help to secure the consistent and broad application of safety measures, and should support smaller companies to comply with the Duty of Care.

It seems likely that an emphasis will be required on the technical challenges associated with data sharing, for example the technical parameters of integrating data sets, and exploration of the use of data trusts. This is important to resolve, because without cooperation, there is the risk that only the largest platforms are able to comprehensively tackle some more complicated harms.[105] This could result problems that are not solved, but moved to smaller sites.[106]

In recent months, tech firms have made good progress in developing new mechanisms for collaborating on the child abuse threat. Principally, this includes the formation of a new global body, Project Protect, that broadly mirrors the existing arrangements for counter terrorism (GIFCT).

International cooperation is welcome, but it will be important that the regulator's focus does not become overly determined by it. Platforms may well argue that UK regulatory expectations should be heavily influenced by the global objectives of Project Protect, and by the business plan it sets out.

However, the regulator should remain focussed on its own regulatory parameters.

The regulator must remain live to both the benefits and risks of industry cooperation. This includes the risk that platforms 'look responsive simply through the performative act of working together, and creating institutional auspices for their actions'.[107] Ofcom must therefore retain a clear focus on outcomes.

Platforms might also be minded to develop an unhelpful 'pack approach', which could dilute regulatory criticism, and create consensus around a problem and its dimensions that may be advantageous from an industry perspective. The regulator must therefore ensure it has highly effective mechanisms for external challenge.

101 For example, Google states that while 'we may review content to determine whether it is illegal or violates our Program Policies [...] that does not necessarily mean that we review content, so please don't assume that we do.'

102 It would also reflect best practice in respect of offline child safeguarding, where reviews into safeguarding failures such as the Laming Report highlighted the impact of poor coordination and led to arrangements for Serious Case Reviews to ensure cross-sector learnings.

103 The Age Appropriate Design Code is clear that sharing data for safeguarding purposes is a 'compelling reason' to process data.

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

105 Alex Stamos, of Stanford University and formerly Facebook, has argued that the 'long tail of social platforms will struggle with [online harms] unless there are mechanisms for the smaller companies to benefit from the research the large companies can afford'.

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

107 Ibid.

NSPCC

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

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

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

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

nspcc.org.uk

EVERY CHILDHOOD IS WORTH FIGHTING FOR