Why the Online Safety Bill fails, and what can make the internet safer

verdict.co.uk/why-the-online-safety-bill-fails-and-what-can-make-the-internet-safer

September 9, 2021

The Online Safety Bill is a proposed Act by the UK Government to protect children online and remove website and social content harmful to both children and adults. If the draft Bill makes it through Parliament, networks failing to comply will face penalties of up to £18 million, or 10% of annual global turnover. The government-approved communications regulatory body Ofcom will also get powers to block offending websites.

Published as a draft in May of this year, the Bill has been raising controversy since its earliest form as 2019's Online Harms White Paper. Critics say the Bill in its current form is too vague in its wording, poses a threat to freedom of expression and places too much power in the hands of social networks. Its supporters meanwhile argue it could be the silver bullet needed to fight online trolls, illegal pornography and some forms of online fraud. They also point to a clause which repeals the government's abandoned internet age verification scheme, although it is likely age verification will be needed to access sites such as .

As GlobalData recently explained in a report on the social media landscape, this regulatory pressure comes from an emerging consensus "that governments should hold social media companies responsible for the content they publish, as it can encourage anti-social and criminal behaviour.

"By requiring social media companies to remove illegal speech swiftly, such laws will reduce online misinformation," the report continues. "However, they also bypass important due process measures and incentivize social media companies to censor content rather than risk a fine."

An obvious fine line exists in protecting people from misinformation and online harm whilst protecting free speech, one which the Online Safety Bill arguably does not tread carefully. Social media brands are also failing in their duty, with automated filter systems failing to moderate content, and the mental health pressures on human moderators of serious concern.

So what is the solution? To find out, *Verdict* explores the Bill with Cris Pikes, CEO and co-founder of Image Analyzer, Geraint Lloyd-Taylor, partner at law firm Lewis Silkin, Peter Lewin, senior associate at Wiggin LLP, Yuval Ben-Itzhak, chief of strategy at Emplifi, David Emm, principal security researcher of Kaspersky, Yuelin Li, VP strategy at Onfido and Paul Bischoff, privacy advocate at Comparitech.

Through separate discussions with these experts we investigate the Online Safety Bill's problems, what it gets right and how the internet can be made a safer space for users without compromising on fundamental rights.

Giacomo Lee: What changes will the Online Safety Bill bring, exactly?

Cris Pikes, Image Analyzer: A lot of people think that the Online Safety Bill only applies to the social media giants and that if, for example, they are running a website where people can upload their holiday snaps and videos and comment on each other's posts, it won't apply to them. This is not the case. The Online Safety Bill will make all digital platform operators responsible for swiftly removing illegal and harmful content to prevent users' posts from harming other people, particularly children.

Online content moderation is akin to operating a swimming pool. Whether you're running an Olympic-sized sports facility, or a paddling pool, you are responsible for keeping that environment healthy and safe for all users. All interactive website operators would be wise to read the Bill to understand their responsibilities and make the necessary preparations to remain on the right side of the law.

> Whether you're running an Olympic-sized sports facility, or a paddling pool, you are responsible for keeping that environment healthy and safe for all users. All interactive website operators would be wise to read the Bill to understand their responsibilities.

Peter Lewin, Wiggin LLP: The Bill is not the only new set of rules concerning online safety that businesses will need to grapple with. The Age Appropriate Design Code (aka the Children's Code) is a new set of guidance from the ICO (the UK's data protection authority) that requires online services that are "likely to be accessed by children" (children being anyone under 18) to ensure that their services are age-appropriate and that steps are taken to mitigate various types of potential harms (financial, physical, emotional etc).

The Code will be enforced from September 2021 and will undoubtedly overlap with proposed aspects of the Online Safety Bill, so it remains to be seen how the Code (enforced by the Information Commissioner's Office, the ICO) will work alongside the Online Safety Bill (enforced by Ofcom) in practice. Several other countries are also considering similar issues and legislation of their own, which may introduce further compliance headaches for global online businesses.

Paul Bischoff, Comparitech: A clear definition of "harm" has not been decided yet, so how this enforcement plays out and what content it affects remains to be seen. The bill could have a similar effect to the proposed repeal of Section 230 of the Communications Decency Act in the USA, which protects tech companies from legal liability for content posted by users. As it stands, much of the bill uses vague language that could threaten freedom of speech.

Service providers will most likely be tasked with removing content and accounts deemed harmful. It's not clear whether this will require tech companies to pre-screen content, which has huge implications for online free speech, or whether harmful content just needs to be removed after it's been posted, perhaps before reaching a certain number of users. The latter is pretty much what tech companies have been doing up to this point anyway and might not have much of a material effect on online harm.

There are a few problems with this approach: It makes private US companies gatekeepers for online content. The very companies that the UK is trying to reign in become the arbiters of what speech should be allowed online. It (also) attacks the messenger, punishing social media companies for content posted by users, instead of going after the real perpetrators of harmful content.

> The very companies that the UK is trying to reign in become the arbiters of what speech should be allowed online.

App stores might (also) be required to remove apps deemed harmful from the UK version of their storefronts.

What is your view on the Online Safety Bill?

Geraint Lloyd-Taylor, Lewis Silkin: There is a real risk that the current Bill creates a two tier system where journalists enjoy extensive protections around what they can say on social media, while ordinary citizens face censorship: ordinary people should not be treated as "second class citizens" in this way.

Yuval Ben-Itzhak, Emplifi *(pictured below)*: Over the last few years we've seen the major platforms really doubling down on removing digital pollution from their online environments. They are doing this in the interest of advertisers, of users, but most of all, in their own interest. They want to make sure their platforms remain appealing over time.

Digital marketing holds significant potential for brands, allowing them to reach and engage with their target audiences. But, in today's world, nothing is more important than brand reputation, purpose and ethics. Brands want to be sure they are choosing safe and trustworthy platforms, free from harm and toxicity, to interact with their customers on and invest their ad spend into.

Lloyd-Taylor: The question of paid-for content will need to also be considered carefully. It is likely that paid-for advertisements will be dealt with separately, still falling within the remit of the Advertising Standards Agency (ASA), with the Competition and Markets Authority (CMA) and other regulators as a backstop, and it makes sense in many ways to exclude these content types from the Bill. It is, after all, aimed predominately at protecting social media users from other individual users, as well as terrorists and rogue actors.

It is interesting that, in theory at least, it is relatively easy for individuals to take out advertisements and publish their thoughts and comments in paid-for space on social media. Thought will need to be given to this issue lacuna.

> It is relatively easy for individuals to take out advertisements and publish their thoughts and comments in paid-for space on social media.

David Emm, Kaspersky: Although the Bill outlines a requirement for platforms to remove "priority illegal content", such as romance scams, this requirement only governs user-generated content, meaning that there's nothing to stop threat actors using advertising on these platforms as a means to defraud people.

What is the business view on the Bill?

Lewin: Some savvy businesses will undoubtedly try and turn the Online Safety Bill to their advantage by championing how "safe" their services are compared to those of their competitors. However, for most businesses, the compliance costs will likely far outweigh any such benefits.

The draft Bill is incredibly complex and large and important aspects of it are still unknown (e.g. aspects which will be set out later under secondary legislation and Ofcom codes and guidance). As a result, businesses will likely spend many months if not years simply getting to grips with their potential new obligations, let alone start implementing the necessary technical and procedural changes.

Businesses are hopeful that the upcoming months of scrutiny and debate will bring some much-needed clarity to these core issues, which may help assuage at least some of these complaints.

> Businesses will likely spend many months if not years simply getting to grips with their potential new obligations.

What is a better alternative to the Bill?

Ben-Itzhak: We should also consider a model of shared responsibility. Applying this duty of care to social media platforms, governments, regulators and users would impart a sense of accountability on to all parties, for what remains a significant and complex problem to eradicate.

Bischoff: Instead of mandating that private social media companies act as the government's gatekeepers to free speech, I think we need to go after actual perpetrators. If someone posts something illegal, police should take steps to identify and make an arrest. Libel, slander, incitement and fraud are all already illegal, we just rarely prosecute those crimes. It just seems easier to blame tech companies and censor speech for everyone.

When it comes to child abuse, most abusers will obviously try to hide their identities. The Online Safety Bill doesn't do anything that requires users to verify their identities. It puts the onus of moderation on profit-driven tech companies while doing nothing to hold actual users accountable. I think social media and tech companies should require identity verification of some kind for users who start or moderate Pages, Groups and chat groups above a certain number of people, for example.

> The Online Safety Bill doesn't do anything that requires users to verify their identities. It puts the onus of moderation on profit-driven tech companies while doing nothing to hold actual users accountable.

Emm: In the absence of a written constitution to provide further checks and balances it seems far more sensible to limit this legislation to strictly unlawful content: compelling platforms to take action in relation to content that is unlawful, and to re-think the unhelpfully vague concept of harms. Parliament could continue to legislate as necessary to make other very harmful activities "unlawful" in order that they are also caught.

That is Parliament's role, and it seems infinitely preferable for Parliament to debate and legislate on specific issues, rather than relying on the combined efforts of the platforms, Ofcom and the government to take these decisions based on a subjective interpretation of harms ad hoc.

Yuelin Li, Onfido *(pictured below)*: We need to find a solution that supports the positive uses for anonymity, while preventing others from abusing anonymity to target hatred and vitriol online without recourse.

Depending on how social media platforms want to embrace identity verification, or indeed retain some level of anonymity, an investment in a digital identity infrastructure would be required (such as an app managed by the user) for users to share only the information that is necessary with the platforms. This allows the creation of different tiers of accounts, from fully anonymous, to real person, to verified real person. Each user can then decide what level of access they want to the platform using these classifications.

> An investment in a digital identity infrastructure would be required… for users to share only the information that is necessary with the platforms.

Will the Bill go the same way as the government's abandoned plans for an online age verification system?

Bischoff: The age verification system was a privacy nightmare and required consent from both users and compliance from porn sites. It was impossible to enforce and no one wanted to share their porn-viewing habits with the government.

The Online Safety Bill only requires compliance from tech companies, so it could have a more material impact and face less backlash from the public. It could result in some tech companies exiting the UK or censoring far more content for UK users.

Is the task of moderating content too Herculean for social giants?

Bischoff: The volume of user-generated content is too high to all be moderated by humans, and automated filter systems don't catch everything.

Ben-Itzhak: Passing this legislation will undoubtedly force platforms to double down on their efforts to limit harmful posts. One thing we know for sure is that no digital platform is flawless. Whether it's an operating system, a mobile device or a network security product, it is loaded with vulnerabilities that can be taken advantage of to harm businesses and individuals.

> No digital platform is flawless.

Li: Governments are putting more pressure on social media platforms to be much faster when responding, moderating and reporting abusive content. Although they are starting to work more closely together, platforms must also be willing to threaten timeouts or bans to have an impact.

Find GlobalData’s Thematic Research Social Media report here.