The UK's Online Safety Bill: The day we took a stand against serious online harms or the day we lost our freedoms to platforms and the state?

Received: 25th April, 2022

Alexander Dittel
Partner in Technology, Wedlake Bell, UK

Alexander Dittel is Partner in Technology for independent UK-based law firm, Wedlake Bell.

Wedlake Bell LLP, 71 Queen Victoria St, London EC4V 4AY
E-mail: adittel@wedlakebell.com

Abstract This paper discusses the UK's Online Safety Bill, which is intended to protect vulnerable individuals online, although at the risk of promoting surveillance techniques and mandating proactive content removal by platforms. It analyses how the Bill, a very ambitious project, tries to safeguard vulnerable people through means which could be easily abused, and asks whether the risk of abuse that could affect everyone is worth the protection of a minority of online users. Recently demonstrated authoritarian approaches to solving the COVID-19 crisis make this concern palpable. The paper concludes by saying that once we take a path, it will be difficult to walk it back.

KEYWORDS: online harms, Online Safety Bill, lawful but harmful, user-generated content, content monitoring, user, monitoring, cyber offences

INTRODUCTION

It is likely that we have all encountered online content that we hope would never be seen by children and vulnerable people. Harmful content has the potential to significantly affect children and members of minority groups and the LGBT community, but also shoppers, readers, voters, patients and others. Online harms are real and the UK Government has spent the last five years working on the Online Safety Bill (hereafter, Bill),[1] which proposes that platforms such as Facebook, Google and Twitter, as well as many small and medium-sized businesses, should have a duty to identify and take action against harmful content. The EU's proposed Digital Services Act has similar aims but resists imposing general monitoring duties.

A decade ago it would have been unthinkable to impose a general obligation on platforms to monitor content. Such an obligation would be inconsistent with freedom of speech and it would place a disproportionate burden on platforms that could not comply without employing significant staff and resources and making limiting changes to their services. Today, the UK Government is not discouraged by these arguments. Machine learning and other technologies have matured to help tackle the growing online harms with sentiment analysis and predictive policing.[2] Nowadays, most platforms operate reactive

and proactive content moderation processes. The Bill aims to improve UK citizens' online safety by moderating, monitoring and removing illegal and harmful content. Under the Bill, platforms are encouraged to deploy profiling and behaviour tracking to protect their users. These duties will not only affect content but also the way in which it is displayed and recommended to users. The finer detail about how platforms can fulfil their new responsibilities and discretions will be outlined by a mandatory code of conduct, which will be issued by the UK's Office of Communications (Ofcom).

While TV content is subject to Ofcom's Broadcasting Code, there is currently no such regulation for user-generated content (UGC) online, which is arguably consumed at a much larger scale. The UK Government claims that the Bill will restore the rule that 'what is illegal offline is illegal online'. The EU Commission makes the same claim about its Digital Services Act. However, in fact, the Bill creates novel duties that have no equivalent offline. And perhaps rightly so.

It might not be appropriate to draw analogies from the offline world, where often the police fail to take action on hate speech, which in one of the worst cases, escalated to the victim being beaten unconscious and his body being set alight.[3] The large-scale distribution of harmful content by unknown perpetrators that is enabled by the digital world cannot be tackled in conventional ways. Despite occasional success stories,[4] the fact remains that the volume of hate speech is enormous and the police do not have the resources or the capacity to investigate all the complaints that are made.[5]

The Bill is expected to impact 24,000 platforms, including 'user-to-user' content sharing service providers, such as social media platforms, discussion forums, gaming platforms and private messaging services; and 'search service' providers, such as global search engines for the web and databases. In addition, pornography websites and fraudulent online adverts are also within its scope. Exempt from the reach of the Bill are services including email, SMS, MMS and user review only services, including news publishers and internal business, public, education and childcare services. Search services that only search one website or database are also excluded. Thankfully for some, the safety duties under the Bill are subject to 'proportionality', which takes account of smaller platforms' limited resources.

Those opposed to the Bill raise concerns about Ofcom's ability to set and enforce rules in this complex area. The new power of platforms to prevent content from being seen could interfere with our fundamental rights. At a recent debate, a Member of Parliament shared with the audience her personal experience of how a death threat received by letter can be prosecuted within months while a death threat made online may take years to prosecute. It is argued that the proposed Bill could deprive us of evidence against perpetrators of criminal and civil offences. To increase accountability, the Bill proposes that adult users will be given the option not to interact with unverified users.

Restricting content on the basis of a presumption of illegality or harmfulness is a level of prescription by the state that is unheard of in modern democracies. A recent paper[6] of the European Parliament discusses if 'burdening Ops [online platforms] with liabilities and high sanctions against the diffusion of extremist content of their services, raises serious risks of over-detection and over-removal, possibly leading to an unacceptable restriction of users' rights and freedoms'. However, the UK Government is determined to 'fight for a new digital age which is safer for users'.[7] The Bill proposes that platforms and search engines will pay Ofcom an annual fee to regulate calculated from their worldwide revenue.

WHAT ARE ONLINE HARMS?

The definition of online harms is essential because it will influence which content might be subject to the duties imposed under the Bill. The latest draft of the Bill offers a broad definition of harms, including any physical or psychological harm or the risk of such harm.

There is no static list of harms and the remit has widened since the Government's initial White Paper in 2019,[8] which presented the harms as given in Table 1.

The duties under the Bill are not limited to actual harms. Harm will include situations where, as a result of the content, an individual acts in a way that increases the likelihood of harm to themselves or does or says something to another person that increases the likelihood of harm to them. It will also include any harm that may arise from its nature or manner of dissemination, such as content being repeatedly shared by individuals.

In relation to 'illegal content', any harm that gives rise to a risk of a criminal offence could be in scope. As a result, any criminal offence, for example economic offences such as money laundering or fraud, could fall under the scope of 'harm' under the Bill.

The latest draft of the Bill promises that Ofcom's code of practice will be voted on by the UK Parliament, which is expected to provide sufficient certainty and ensure that platforms cannot be incentivised to over-remove legal material as a result of taking a wider interpretation of harm than intended.

FILLING A GAP IN REGULATION?

It is unlikely that individuals vulnerable to online harm would have the mindset and resources to bring a court action against those causing them harm. Often the victims do not know about the online harms they are facing until it is too late, for example, teenagers taking a substance that is wrongfully promoted online by multiple parties as having aesthetic effects on their body. It is difficult to see how one could effectively protect these vulnerable individuals without some kind of intervention. Historically, platforms have been reluctant to remove content, citing freedom of speech, the public interest of users in receiving information and net neutrality.

Is there a commercial side to harmful content? Driving traffic through popular

Table 1: Online harms as presented in the Government's Online Harms White Paper consultation.

Harms with a clear definition	Harms with a less clear definition	Underage exposure to legal content
• Child sexual exploitation and abuse. • Terrorist content and activity. • Organised immigration crime. • Modern slavery. • Extreme pornography. • Revenge pornography. • Harassment and cyberstalking. • Hate crime. • Encouraging or assisting suicide. • Incitement of violence. • Sale of illegal goods/services, such as drugs and weapons (on the open internet). • Content illegally uploaded from prisons. • Sexting of indecent images by under 18s (creating, possessing, copying or distributing indecent or sexual images of children and young people under the age of 18).	• Cyberbullying and trolling. • Extremist content and activity. • Coercive behaviour. • Intimidation. • Disinformation. • Violent content. • Advocacy of self-harm. • Promotion of Female Genital Mutilation (FGM).	• Children accessing pornography. • Children accessing inappropriate material (including under 13s using social media and under 18s using dating apps). • Excessive screen time.

Source: https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper

content, whether deemed good or bad, allows platforms to monetise content. However, developments such as large-scale copyright infringement, terrorism, cyberbullying, online fraud, fake news, misinformation about vaccines and interference with the democratic process[9] have pushed platforms to adopt reactive and proactive measures in content management.

Some of these changes were pushed by legislation, such as the Digital Millennium Copyright Act (DCMA) in the US or defamation laws in the UK. However, in the majority of cases, online harms remain unregulated and subject to voluntary initiatives focused on the protection of children and vulnerable groups. For example, some charities that focus on the protection of children online work directly with Facebook to take down specific content. On the other hand, members of the LGBT community complain about platforms' inaction following compliant submissions.[10] Platforms apply their own community standards and terms of use and therefore enjoy a wide discretion in moderating content. YouTube, for example, reserves the right to remove any content that may cause harm to others. Platforms have, however, been criticised for failing to follow their own community standards and terms of service.

In the UK, platforms enjoy immunity from liability for unlawful content that they are unaware of under the mere conduit and hosting exemptions provided in The Electronic Commerce (EC Directive) Regulations 2002, as amended. Similar immunity is granted under Section 230 of the US Communications Decency Act of 1996. While there is no positive duty to remove unlawful content, upon notice, platforms must act expeditiously to remove or to disable access to the information or face liability.

The traditionally reactive approach to act upon notice is now combined with focused proactive monitoring. However, given that curation of content could invalidate the immunity and interfere with fundamental rights, these initiatives come rather slowly. While unlawful content would likely be removed, the question about content that is not unlawful but still harmful is entirely at the platform's discretion. Platforms use a combination of algorithms and human reviewers to flag and remove content. However, this power in the hands of platforms has led to a high number of 'false-positives', for example, where journalistic content is mistakenly removed for alleged breach of community standards.

WHAT ARE THE CORE DUTIES UNDER THE BILL?

The level of duties will depend on the provider's activities and size, as follows:

- All providers will have duties in relation to illegal content. This will include a duty to assess risk and to take proportionate protective measures, depending on the seriousness of the illegal content. Providers will also have content reporting and record keeping duties; will have to operate a complaints procedure to deal with improper tracking, content removals, delisting or reduction in ranking; and implement measures to safeguard freedom of expression and privacy.
- Services likely to be accessed by children will have an additional duty to assess risk and take protective measures in relation to content that is harmful to children, even if such content is legal. They will also have to report new non-designated harmful content to Ofcom.
- All 'category 1 providers' with a large volume of users and significant functionalities, including the likes of Facebook, Google and Twitter, will have an additional duty to assess risk and take measures to protect adult users against content that is harmful, even if such content is legal. They will also have duties

to empower adult users by providing control over content, protect content of democratic importance, protect journalistic content, safeguard freedom of information and privacy, and report new non-designated harmful content to Ofcom.

The Bill imposes additional duties in relation to specific areas of concern:

- All platforms will have a duty to report any child sexual exploitation content to the National Crime Agency.
- Large providers must deploy proactive technologies and use proportionate systems designed to minimise the occurrence of fraudulent advertising.
- Publishers of pornography must ensure that children are not normally able to encounter such content.
- Upon notice by Ofcom, all platforms must use accredited technology to identify and take down terrorism content or child sexual exploitation content, whether communicated publicly or privately.

THE DAWN OF GENERAL MONITORING OBLIGATIONS?

The Bill is not the only example of how advances in technologies such as machine learning are starting to shape the law.

While the E-Commerce Directive (Directive 2000/31/EC) prohibits imposing a general monitoring obligation on services providers, it does not preclude a monitoring obligation relating to a specific case. In the case of Eva Glawischnig-Piesczek v Facebook Ireland Limited Case C-18/18, concerning the right to be forgotten, the Court of Justice of the European Union (CJEU) held that a national court can legitimately request the host provider, as part of a specific monitoring obligation, to remove any content identical to the offending content but also any equivalent content. The injunction would have to specify all kinds of word combinations that are in the scope of equivalent content. The CJEU balanced 'effectively protecting a person's reputation and honour' under an injunction with not 'imposing an excessive obligation on the host provider'. Essentially, the host provider must not be required to carry out an independent assessment of every piece of content.

However, the CJEU struggled with this limitation because illegality stems from defamatory statements and not from a combination of words. Facebook's recourse to automated search tools and technologies allowed it to carry out the search without manual review of content. The available technology influenced the CJEU's decision.

Another example is the Age Appropriate Design Code (Children's Code) issued by the UK Information Commissioner's Office (ICO) in 2020. The Code anticipates that with the emerging age assurance technologies, controllers will be better equipped to comply. Extensive user profiling and behaviour analysis will help identify children and enable controllers to act in the best interest of the child by restricting harmful data processing.

While the Bill will impose a general monitoring obligation, the EU's draft Digital Services Act[11] preserves a prohibition on a general monitoring duty.

WHAT ARE THE PROPOSED SAFETY DUTIES OF PLATFORMS?

The Bill's safety duties impose an obligation to identify and mitigate the risk of harm.

Illegal content

In relation to illegal content, platforms will have a duty to:

- carry out an assessment of risks of harm to individuals that may arise from illegal content on the platform;
- take or use proportionate measures to effectively mitigate and manage those risks;

- operate a take-down process for the swift removal of illegal content upon notice; and
- implement proportionate systems and processes designed to prevent individuals from encountering and minimising the presence of priority illegal content. According to a press release by the Department for Digital, Culture, Media and Sport,[12] a list will be set out in the Bill and will include revenge porn, hate crime, fraud, the promotion or facilitation of suicide, people smuggling and sexual exploitation.

The duties have attracted much criticism. Most people will agree that illegal content should not be present on platforms, but the question is who will decide whether content is legal? Platforms are of course best placed to monitor their content. However, the Bill will give each platform the power to make decisions about content based on its assessment that the content amounts to a relevant offence. Ofcom will issue a binding code of practice regarding how to go about this, which must be laid before Parliament.

Some of the safeguards to preserve fundamental rights include the requirement for platforms to set out in their terms of service how individuals will be protected from illegal content and a requirement to apply those provisions consistently. Individuals must be informed of their right to bring action if content is restricted unlawfully. Platforms will have a duty to balance protecting freedom of expression and guarding against breaches of data protection law. Category 1 platforms will have to carry out an impact assessment of their measures on such content.

Protecting children and adults from harmful content (whether lawful or not)

In relation to protecting children, platforms likely to be accessed by children will have a duty to:

- carry out an assessment of risks to children;
- take or use proportionate measures to effectively mitigate and manage those risks;
- mitigate the impact of harm to children in different age groups presented by content that is harmful to children;
- implement proportionate systems and processes designed to prevent children of any age from encountering primary priority content designated as such by the UK Government; and
- implement age assurance and similar techniques to protect children in age groups judged to be at risk of harm from encountering a particular kind of content, such as pornographic content.

Apart from Ofcom's code of practice on how to comply, the UK Government will issue regulations to designate priority content that is harmful to children. However, the duty applies regardless of such designation to content which the platform believes presents a material risk of significant harm to an appreciable number of children in the UK.

In relation to protecting adults, category 1 platforms will have a duty to:

- carry out an assessment of risks to adults;
- summarise their findings in its terms of service;
- specify in their terms of service how each type of priority content that is harmful to adults will be blocked, restricted or subject to limitations on promotion and recommendations; and
- prevent, minimise and disable any fraudulent adverts.

The duties in relation to harmful but lawful content have attracted the most debate due to the difficulty of defining the threshold that would trigger the duty and the potential implications of automated content removals. Platforms will act on the reasonable belief that content may cause harm. However, the scope of the duties is guided by the platform's risk assessment and Ofcom's codes

of practice and risk profile designations. Ofcom will have a difficult task in providing the right guidance on what content could constitute harm to the fictional audience member of reasonable sensibilities.

Protecting journalistic and democratic content

News publisher content is excluded from most of the duties under the Bill. In addition, in relation to content of democratic importance, platforms must ensure their content and user monitoring and restriction systems and processes are designed 'to ensure that the importance of the free expression of content of democratic importance is taken into account' regardless of political opinion. A similar duty applies to the 'free expression of journalistic content'.

This creates an opposite duty to those related to illegal content and harmful content. Instead of a general presumption of freedom of speech, platforms will have to moderate and take down content while at the same time ensuring that any content of democratic importance and journalistic content are not affected. Platforms will be restricted from removing such UGC. It is difficult to envisage this without a comprehensive system of content categorisation and user profiling in place.

A DUTY TO USE PROACTIVE TECHNOLOGY?

The Bill presumes that technology is available to comply with the new duties. The Government suggests that organisations use 'automated or human content moderation, banning illegal search terms, spotting suspicious users and having effective systems in place to prevent banned users opening new accounts'.[13] There is no explicit obligation to use 'proactive technologies'.

Proactive technologies include:

- 'Content moderation technology', including algorithms, keyword matching, image matching or image classification, which analyse relevant content to assess whether it is illegal, harmful to children or fraudulent advertisement.
- 'User profiling technology', which analyses relevant content, user data, or metadata relating to relevant content or user data, for the purposes of building a profile of a user to assess characteristics such as age.
- 'Behaviour identification technology', which analyses relevant content, user data, or metadata relating to relevant content or user data, to assess a user's online behaviour or patterns of online behaviour (for example, to assess whether a user may be involved in, or be the victim of, illegal activity).

There is a fine balance between deploying intrusive tracking and preserving the right to privacy. It is plausible that certain tracking may not be permitted under data protection laws, even if it is the only way to fulfil the intentions of the Bill. However, if the Bill's focus is about moderating content, it could be argued that tracking and profiling of users and making automated decisions about them should not be legally required.

The difficulties in user profiling in order to protect members of the LGBT community was demonstrated in a recent ban on the Shinigami Eyes browser extension in Norway. The tool highlights transphobic and trans-friendly content and individuals based on profiling, giving trans people the confidence to engage online without being offended. The subjective assessment takes place covertly without the ability of the marked person to express his or her views. The marked person could lose their job, friendships and be targeted by hate speech as a result of being marked. The Norwegian regulator found that it could have a chilling effect on the ability and willingness of individuals to participate in online discourse, through fear of receiving a marking and subsequently suffering negative consequences. It was

feared that the tool removed the need for the individual to make their own assessment, which could strengthen the echo chambers found online.

Nevertheless, Ofcom may impose the requirement to implement proactive technologies by way of a so-called 'confirmation decision'. A proactive technology measure may be recommended only for the purpose of compliance with duties in relation to illegal content, content harmful to children and fraudulent advertising. It is expected that Ofcom will issue a Code of practice on proactive technology measures. Another limitation is that proactive technology must not be recommended to analyse UGC communicated privately or the underlying metadata.

RISK ASSESSMENTS

The Bill relies heavily on suitable and sufficient risk assessments to define the scope of the safety duties. Similar to a risk assessment for compliance with the Children's Code, such assessments will have to consider the platform's audience, the risks likely encountered and which measures are in place to mitigate those risks.

A content risk assessment should consider:

- the user base, including the number of children in different age groups;
- the level of risk to users in respect of each kind of priority and non-priority (with each kind separately assessed), and non-designated content, taking into account (in particular) algorithms used by the service, as well as how easily, quickly and widely content may be disseminated by means of the service;
- the level of risk of harm to individuals presented by content of different kinds;
- the level of risk of functionalities of the service facilitating the presence or dissemination of content, and identifying and assessing those functionalities that present higher levels of risk, including the ability of adults to find and contact other users, including children;
- the different ways in which the service is used, and the impact of such use on the level of risk of harm that might be suffered by individuals;
- the nature and severity of the harm that might be suffered by individuals from these matters; and
- how the design and operation of the service (including the business model, governance, use of proactive technology, measures to promote users' media literacy and safe use of the service, and other systems and processes) may reduce or increase the risks identified.

Ofcom will publish its own risk assessments about illegal and harmful content.

THOSE OPPOSED TO THE BILL SAY . . .

A major point of debate is the fact that platforms and their algorithms will make decisions about content based on their duty guided by Ofcom, risk assessments and reasonable grounds. For example, there will be no independent adjudication of illegal content. As commentators point out, there is a real risk that automated removals of content may breach presumption against prior restraint.[14] As expressed by William Blackstone, 'The liberty of the press is indeed essential to the nature of a free state; but this consists in laying no previous restraints upon publications, and not in freedom from censure for criminal matter when published.'[15] There should be no presumption of harm arising from certain speech. Being offensive is not an offence. However, it is difficult to see how platforms can comply with the Bill without such presumptions.

Moderating potentially upsetting content should not be the duty of the state and its agents in a democratic society. If the

Bill is not specific enough about harms, platforms' compliance will be arbitrary. If the threshold for harm is too low, for example, to safeguard those prone to being easily offended, then everyone else could lose out in being denied access to content. There is a concern that it will be impossible for platforms to get it exactly right when implementing their own interpretation of the obligations and for Ofcom to regulate effectively without significant resource, which is not provided for under the Bill.

Some argue that 'Social media service providers should each be seen as responsible for a public space they have created, much as property owners or operators are in the physical world'.[16] Others say that if the Government wishes to consolidate online and offline safety, as is claimed, the duties under the Bill are excessive and go well beyond what we would expect in the offline world.

The platforms' obligation to moderate content could, in practice, often translate into removing it, as platforms already do, if there is a risk that their duty is triggered. Given the backdrop of the heavy penalties and potential criminal liability under the Bill, many platforms may err on the side of caution. Such automatic removal could mean that evidence for criminal proceedings is destroyed and victims may never know they were a target, or that such evidence ever existed. Preventative action by platforms will affect individuals' freedom of access to content and freedom of speech, particularly when content is automatically blocked by imperfect algorithms. In addition, if content is automatically removed, organisations may not be able to comply with the fundamental right to access personal data under data protection laws.

The Bill is much more comprehensive than any similar laws in other countries, which tend to target hate speech.

It is possible that compliance with the Bill will affect smaller content sharing platforms' ability to compete and innovate.

The Bill will introduce mandatory age verification for services that are not intended for children, something that is encouraged under the Children's Code. However, this will also mean that content is not readily accessible and platforms have an incentive to collect more data about their users.

There are concerns that the Bill will effectively ban end-to-end encryption, as duties are impossible to comply with if the platform has no access to content. The Bill's impact is far-reaching as it will not only apply to public content but also private communications, save for those that are excluded.

DOES THE BILL HAVE EXTRA-TERRITORIAL EFFECT?

Yes, it does. In a global world of digital technologies, the offending content could come from anywhere and restricting the Bill's reach would defeat its purpose. In fact, there has been a trend of expanding the geographical reach of laws tackling digital technologies.

In its landmark ruling in Case C-507/17 Google v CNIL, the CJEU held that there is no obligation under EU law for Google to apply the European right to be forgotten under the General Data Protection Regulation (GDPR) globally. The decision clarifies that, while EU residents have the legal right to be forgotten, the right only applies within the borders of the bloc's 28 member states. However, at the same time, the court attempted to establish the lawfulness of global de-referencing as a general principle. By finding that EU law does not prohibit such orders, member states remain able to order search engine operators to de-reference globally after balancing the conflicting rights of personal data protection against the right to freedom of information under national standards of protection of fundamental rights.

In a similar spirit, the Bill applies to all platform services with a link to the UK,

either due to a significant number of UK users or the targeting of UK audiences. However, the scope of the safety duties under the Bill extends only to the design, operation and use of the service in the UK or as it affects UK users.

In addition, a platform service will be in scope if it can be used by a UK user and 'there are reasonable grounds to believe that there is a material risk of significant harm to individuals in the United Kingdom presented by' the UGC present in the service.

Territoriality is also important in determining what is illegal content. In this regard, it is not important if any illegal act occurred in the UK or elsewhere.

Ofcom's power to require the production of documents by an information notice includes the power to require the production of documents held outside the UK. Ofcom also has the power to require the attendance for interview of an individual who is outside the UK.

REMEDIES AND SANCTIONS

As we are heading into a digital future, automated decisions will become a part of our daily lives. The Bill will encourage the use of automated decisions about users and their content. It is critically important that users benefit from strong transparency and remedies against unfair automated decisions.

In this regard, the Bill encourages platforms to increase transparency in their terms of service about how content may be treated. Users will have the right to bring a claim against platforms for breach of contract. The Bill also imposes a duty to operate easy to use complaints procedures that must result in appropriate action being taken by the platform. Users will be able to complain about:

- being affected by illegal content;
- content that might be harmful to children or adults;
- being mistakenly taken for a child and blocked from access;
- a platform's failure to comply with its duties in relation to freedom of expression and others;
- their UGC being taken down or restricted;
- a warning issued to the user because of their UGC; and
- the use of proactive technologies resulting in a lowering of the promotion or ranking of UGC or a breach of service terms.

However, the Bill does not create a new statutory private right of action like that in the GDPR. Having said that, the GDPR right of action might be available where personal data is processed in breach of the GDPR when making a decision about content or harm affecting the individual. We are yet to see how courts might respond to these increasingly complex data protection claims.

In terms of regulatory oversight, Ofcom will have the power to request information and interviews in relation to platforms' compliance. It will have the power to issue provisional notices of contravention, confirmation decisions requiring remediation steps and penalties for failure to comply of up to £18 million or 10 per cent of qualifying global annual turnover, whichever is greater. Ofcom can apply to the court for injunctions and specific performance and service restriction orders.

The Bill introduces a number of criminal offences:

- Harmful communications offence — committed when intentionally sending a message with a real and substantial risk of causing serious distress to the likely audience. This offence will address assisting self-harm and the so-called 'epilepsy trolling' (ie sending flashing images to epilepsy sufferers).
- False communications offence — committed when intentionally sending false information that will cause non-trivial psychological or physical harm

to the likely audience. This offence will address hoax calls.
- Threatening communications offence — committed when sending a threat of death or serious harm intended to cause, or reckless as to whether it would cause, fear that the threat would be carried out.
- In relation to the above points, company directors and officers can be personally liable if the communication is sent with their connivance or is attributable to their neglect.
- Sending photograph or film of genitals if intending to cause alarm, distress or humiliation, or to obtain sexual gratification. This offence addresses cyberflashing.

CONCLUSION

Most people will agree that something has to be done to tackle the risks brought by content on the internet, which can significantly affect vulnerable people. Platforms often fail to act on user reports of hate speech or inappropriate content and the Bill will certainly have a positive impact on improving these reactive processes.

However, the Bill remains a very ambitious project. It tries to safeguard vulnerable people through means which could be easily abused. Is the risk of abuse that could affect everyone worth the protection of a minority of online users? Many would argue that it is not. Recently demonstrated authoritarian approaches to solving the COVID-19 crisis make this concern palpable. Once we take a path, it will be difficult to walk it back.

Without appropriate safeguards, transparency, remedies and a mature regulatory culture, taking action could significantly affect individuals' fundamental rights to information and freedom of expression. Deciding for us which content is right or wrong could lead to dictating what we should watch. That does not feel like freedom for anyone.

The text of the Bill is trying to say all the right things. However, the effect of it will really depend on Ofcom's ability to provide mature guidance and avoid strict enforcement action, which would force providers to always err on the side of caution and over-identify and over-remove content.

As the Bill is making its way through Parliament, we can only hope that the practice adopted by the sector and regulators will carry the spirit of what is intended here, that is, safeguarding vulnerable individuals while preserving the full extent of freedom of speech, access to information and the plurality of opinions.

We can only hope that civil societies will stay vigilant. It is a shame that the Bill does not foresee representative action. Thankfully, privacy advocates do not wait for an invitation.

References and notes

1. 'Online Safety Bill', available at https://publications.parliament.uk/pa/bills/cbill/58-02/0285/210285.pdf (accessed on 22nd April, 2022).
2. Pereira-Kohatsu, J. C., Quijano-Sȟchez, L., Liberatore, F. and Camacho-Collado, M. (2019) 'Detecting and monitoring hate speech in Twitter', *Sensors (Basel)*, Vol. 19, No. 21, 4654, available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6864473/ (accessed on 22nd April, 2022).
3. 'Hate crimes in the UK — the victims' stories', available at https://www.amnesty.org.uk/blogs/ether/hate-crimes-uk-victims-stories (accessed on 22nd April, 2022).
4. A Europol-led campaign targeted over 170 individuals in relation to offences such as dissemination of racist and xenophobic hate speech, calls to violence and incitement to commit offences. See, 'Tackling hate crime across Europe: Second joint action day targets over 170 individuals', available at https://www.europol.europa.eu/media-press/newsroom/news/tackling-hate-crime-across-europe-second-joint-action-day-targets-over-170-individuals (accessed on 22nd April, 2022).
5. 'Harry Miller: Being offensive is not an offence says free speech advocate', available at https://www.bbc.co.uk/news/av/uk-england-lincolnshire-59728794 (accessed on 22nd April, 2022).
6. 'Liability of online platforms', European Parliamentary Research Service, February 2021, available at https://www.europarl.europa.eu/RegData/etudes/STUD/2021/656318/EPRS_STU(2021)656318_EN.pdf (accessed on 22nd April, 2022).

7. 'World-first online safety laws introduced in Parliament', Department for Digital, Culture, Media & Sport, available at https://www.gov.uk/government/news/world-first-online-safety-laws-introduced-in-parliament (accessed on 22nd April, 2022).
8. 'Online Harms White Paper', April 2019, HM Government, Department for Digital, Culture, Media & Sport, available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973939/Online_Harms_White_Paper_V2.pdf (accessed on 22nd April, 2022).
9. 'Russia's Putin authorised pro-Trump "influence" campaign, US intelligence says', available at https://www.bbc.co.uk/news/world-us-canada-56423536 (accessed on 22nd April, 2022).
10. 'Top social media platforms "unsafe" for LGBTQ users, report finds', available at https://www.nbcnews.com/nbc-out/out-news/top-social-media-platforms-unsafe-lgbtq-users-report-finds-rcna889 (accessed on 22nd April, 2022).
11. Proposal for a Regulation of the European Parliament and of the Council on a single market for digital services (Digital Services Act) and amending Directive 2000/31/EC, available at https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN (accessed on 22nd April, 2022).
12. 'Online safety law to be strengthened to stamp out illegal content', 4th February, 2022, available at https://www.gov.uk/government/news/online-safety-law-to-be-strengthened-to-stamp-out-illegal-content#:~:text=Digital%20Secretary%20Nadine%20Dorries%20today,people%20smuggling%20and%20sexual%20exploitation (accessed on 22nd April, 2022).
13. Ref. 12 above.
14. Smith, G. (2022) 'Harm Version 4.0 — the Online Safety Bill in metamorphosis', 19th February, available at https://www.cyberleagle.com/2022/02/harm-version-40-online-safety-bill-in.html (accessed on 22nd April, 2022).
15. 'Right to know v prior restraints', 7th March, 2007, available at https://www.theguardian.com/media/2007/mar/07/pressandpublishing.partyfunding#:~:text=William%20Blackstone%2C%20an%2018th%2Dcentury,for%20criminal%20matter%20when%20published.%22 (accessed on 22nd April, 2022).
16. Woods, L. and Perrin, W. (2019) 'Internet Harm Reduction: A proposal', 30th January, available at https://www.carnegieuktrust.org.uk/blog-posts/internet-harm-reduction-a-proposal/ (accessed on 22nd April, 2022).