iea

REGULATORY AFFAIRS

MIKOŁAJ BARCZENTEWICZ
and MATTHEW LESH

FEBRUARY 2022

IN HARM'S WAY

Why online safety regulation needs an Independent Reviewer

IEA papers are designed to promote discussion of economic issues and the role of markets in solving economic and social problems. As with all IEA publications, the views expressed are those of the author and not those of the Institute (which has no corporate view), its managing trustees, Academic Advisory Council, or other senior staff.

Contents

About the authors

Mikołaj Barczentewicz (@MBarczentewicz) is a Senior Lecturer in Law and the Research Director of the Law and Technology Hub at the University of Surrey. He is also a Senior Scholar at the International Center for Law & Economics, a Research Associate of the University of Oxford Centre for Technology and Global Affairs and a Fellow of the Stanford Law School and University of Vienna Transatlantic Technology Law Forum. His research spans technology law and policy, UK and EU public law, and legal philosophy. Mikołaj is a graduate of the University of Oxford, where he obtained his MJur, MPhil, and DPhil degrees, as well as of the University of Warsaw, where he studied law and philosophy.

Matthew Lesh is the Head of Public Policy at the Institute of Economic Affairs. He regularly appears on television and radio, and has written dozens of opinion and feature pieces for print and online publications such as *The Times*, *The Telegraph* and *The Spectator*. He has provided extensive commentary and written various papers and submissions about the Online Safety Bill. He is also a Fellow of the Adam Smith Institute and Institute of Public Affairs. Matthew graduated with First Class Honours from the University of Melbourne with a Bachelor of Arts (Degree with Honours) and completed a Masters in Public Policy and Administration at the London School of Economics, where he received the Peter Self Prize for Best Overall Result.

This paper builds on written evidence for the Joint Committee on Draft Online Safety Bill submitted by the Adam Smith Institute[1] and by Dr Mikołaj Barczentewicz.[2]

The authors are grateful to the anonymous reviewers of this paper.

1 Written evidence submitted by the Adam Smith Institute (OSB0129) to the Joint Committee on Draft Online Safety Bill's Legislative scrutiny inquiry, 22 September 2021 (https://committees.parliament.uk/writtenevidence/39293/html/).

2 Written evidence submitted by Dr Mikołaj Barczentewicz, Senior Lecturer in Law, University of Surrey (OSB0152) to the Joint Committee on Draft Online Safety Bill's Legislative scrutiny inquiry, 28 September 2021 (https://committees.parliament.uk/writtenevidence/39325/html/).

Summary

- The draft Online Safety Bill presents a significant threat to freedom of speech, privacy and innovation. 'Safety' has been prioritised over freedom. The Bill's proponents wrongly assume it is possible to remove 'bad' content without negatively impacting on the 'good' and that platforms, not users, are responsible for 'harms'.
- The Bill's inclusion of 'legal but harmful' speech – along with defining unlawful speech as any content that the platform merely has 'reasonable grounds to believe' is unlawful – risks state-mandated automated censorship of lawful online speech. The duties to 'have regard' to freedom of expression and privacy are far weaker than the 'safety' duties.
- The Bill threatens innovation and competition within the UK economy by imposing byzantine duties that will inevitably be harder and more costly for start-ups and smaller companies to comply with, while discouraging companies from operating in the UK, limiting access to online services.
- The Bill provides extraordinary discretion to the Secretary of State and Ofcom to design 'codes of conduct' that will define 'legal but harmful' content. They will also have the power to impose additional requirements such as age verification and undermine end-to-end encryption. The regulator will also have significant leeway about what types of content and which platforms to target.
- If the Government is unwilling to fundamentally rewrite the Bill, there is a clear need for serious, independent scrutiny mechanisms to prevent regulatory and ministerial overreach.
- An Independent Reviewer of Online Safety Legislation, modelled partly on the Independent Reviewer of Terrorism Legislation, could provide some accountability.

- The Independent Reviewer would need to be properly resourced and empowered to scrutinise the activities of the Secretary of State and Ofcom and communicate findings to policymakers and the general public.
- An Independent Reviewer, properly empowered and resourced, could stand up for freedom of expression, privacy and innovation while being a bulwark against future authoritarian demands.

Introduction

The draft Online Safety Bill ('the Bill') substantially reimagines the role of the state with respect to 'safety' from all forms of speech, whether lawful or not, that could cause any manner of harm, including psychological.

The only way to ensure the Internet is entirely 'safe' is for the Internet to be abolished. Perfect safety is neither an attainable nor desirable goal. It is not desirable because being 'safe', particularly from ideas with which one disagrees, weakens our ability to debate controversial issues and increases the chances that bad ideas are not challenged.

The Bill treats the online world as an environment that could and should be 'child-proofed' to a greater extent than the offline world. In the offline world, we expect adults to know how to use streets and cars in ways that minimise the risk of harm. We also expect adult guardians to instruct and oversee children using public spaces.

Under the Bill, 'child-proofing' will apply very broadly given the low threshold for whether a service is 'likely to be accessed by children'. Even the most invasive age-verification techniques will be highly unlikely to stop motivated under-18s from accessing services not directed at them. An unintended consequence of the Bill may be in motivating many young people to become proficient hackers.

The Bill fits well in the global trend of laws that 'proceed on the false assumption that platforms could remove the bad without the good, and faster, if only they just tried harder' (Douek 2021: 813).

The 'duty of care' model creates a requirement on companies to protect their users. But on platforms with user-generated content, the 'harms' are caused by users, not by the platforms themselves. Even algorithms that

promote content that some consider to be harmful are, at a fundamental level, a reflection of the users' desires. The extent to which social media platforms are promoting harmful content is proportional to the extent to which its users are creating and sharing the content. That means that, in practice, the Bill is not simply 'regulating Big Tech'. It is regulating the legal speech of tens of millions of citizens who use the Internet every single day.

According to the Government, protection of the freedom of expression is one of the three key principles of the Bill.[3] However, the Bill explicitly prioritises safety duties and would in practice require censorship of speech that would be lawful offline. Instead of protecting the freedom of expression through adequate institutional safeguards, the Bill tries to sidestep the issue by vaguely instructing private companies to solve the problem of balancing safety and other values like free expression, but with overwhelming preference given to safety. Furthermore, the Bill does not address the risks to freedom of expression and to innovation that will stem from Ofcom's enforcement and from present and future governments' actions exercising powers under the Bill.

The Bill also threatens innovation and competition within UK economy. The duties imposed on businesses of all sizes (e.g. duties to protect children's safety online, undertake various safety assessments) will create an additional advantage for the biggest players, who will be able to shoulder the costs more easily. Even with greater burdens on larger Category 1 services, many smaller businesses and start-ups could be crippled by the broader compliance costs.[4] The Bill will, for example, require any company seeking to operate in the UK to undertake assessment of potential safety risks before beginning a service. Vague provisions that businesses only need to take 'proportionate' measures are not an adequate answer, especially if it will require costly legal advice to assess what measures are 'proportionate'.

The Government has failed to justify that the Bill is necessary or that the restrictions on our freedoms and on our digital economy are proportionate.

---

3 Written evidence submitted by the Department for Digital, Culture, Media and Sport and the Home Office (OSB0011) to the Joint Committee on Draft Online Safety Bill's Legislative scrutiny inquiry (https://committees.parliament.uk/writtenevidence/38883/html/ at [4], [43]).

4 Category 1 services are the largest user-generated platforms, to be determined by Ofcom and expected to be the likes of Facebook, Twitter and Google, and will be subject to additional duties in relation to content that is harmful to adults (but legal), democratic content and journalistic content.

It may be best to scrap the Bill entirely and pursue meaningful solutions to online issues that actually address serious crime while protecting fundamental freedoms and innovation. However, it is likely that the Bill will proceed in some form. Therefore, this paper not only identifies the main risks in the Bill, but also proposes institutional safeguards that could be included in the Bill.

The proposed safeguards will not entirely prevent infringement of fundamental freedoms or harm to the economy. They will, however, reduce some of the risks, increase regulatory accountability and help ensure that the public is informed about how the new online safety legislation is enforced.

Specifically, this paper proposes the creation of a new position of the Independent Reviewer of Online Safety Legislation, modelled partly on the Independent Reviewer of Terrorism Legislation. The Bill at present lacks sufficient incentives to ensure that Ofcom will protect privacy, freedom of speech and freedom of association, as well as promote useful innovation, while exercising its new, dangerous powers. Moreover, no one – other than perhaps the authors of the Bill – believes that parliamentary and governmental oversight will be effective (e.g. in protecting lawful but unpopular speech). There is no one to 'watch the watchers'. Instead, there is a significant risk that the regulator, armed with substantial new powers, will follow popular demands or institutional groupthink to limit freedom of expression. This paper proposes that a Reviewer should step into that role, scrutinising how the Government and Ofcom will enforce the Bill and continually informing the public about Ofcom's actions that hinder privacy, innovation or freedom of expression.

Risks in the Online Safety Bill

The draft Online Safety Bill suffers from byzantine complexity. There is a serious lack of clarity and specificity in its key provisions. The Bill leaves substantial details to be determined later, by Government and to a large extent by Ofcom (both in respect of the formulation of the all-important codes of practice and, crucially, in the implementation and enforcement of the ensuing regulatory regime). Those details will make the difference between the Bill resulting in performative compliance, perhaps with some beneficial effects, and the Bill transforming the British Internet into an Orwellian public–private censorship and surveillance partnership. The defenders of the Bill assume that the provisions *will* be applied sensibly. The problem is, however, that the Bill provides no guarantee that this will happen. It is also presumed that Ofcom will be perfectly resourced, knowledgeable, and capable of balancing various competing demands within the Bill to achieve something – perfect 'safety' without infringing on liberty – that has proven historically impossible in every context. In sum, policymakers have apparently forgotten about the existence of unintended consequences and trade-offs.

This section summarises the key risks stemming from the Bill in two spheres: risks to fundamental freedoms (freedom of expression, freedom of association and privacy) and risks to innovation and competition.

Risks to freedom of expression, freedom of association and privacy

Inadequate safeguards

The Government claims that 'the Online Safety Bill does not require service providers to remove any legal content'.[5] Leading Internet lawyer Graham Smith has demonstrated that this is misleading. Service providers are set to become 'proxies for the regulator'.[6]

In particular, the Bill would mandate that Category 1 (large) services, likely to include Google, Facebook and Twitter, undertake detailed risk assessments which would identify the (supposed) risks arising from – among other things – lawful but (supposedly) harmful content. There would then be a further 'safety duty' which would require such services to 'take proportionate steps to mitigate and effectively manage' the risks so identified. It is easy to see how that would be interpreted as a requirement to remove legal content.

Moreover, the Bill does not define illegal content by reference to what is or is not an offence, but rather, as any content that the provider has '*reasonable grounds to believe*' (emphasis added) amounts to a relevant offence. By defining 'illegal content' in this way, the draft Bill deems content to be illegal (and thus subject to the safety duties, including removal) that is only arguably illegal. That will inevitably include legal content.

Therefore, the duties imposed by the Bill will result in the removal of legal content. While removing or shadow banning (i.e. hiding posts without informing the author) content, service providers will be acting under the codes of conduct and threat of enforcement action from Ofcom. To assess whether providers are acting in accordance with their legal duties, Ofcom will, in practice, need to evaluate individual pieces of content and decide what is illegal or harmful and what is not.

5 Written evidence submitted by the Department for Digital, Culture, Media and Sport and the Home Office (OSB0011**)** to the Joint Committee on Draft Online Safety Bill's Legislative scrutiny inquiry (https://committees.parliament.uk/writtenevidence/38883/html/ [44]).

6 The draft Online Safety Bill: systemic or content-focused? *Cyberleagle*, 1 November 2021 (https://www.cyberleagle.com/2021/11/the-draft-online-safety-bill-systemic.html).

In clause 12, the Bill imposes on service providers a duty

> to have regard to the importance of —
>
> (a) protecting users' right to freedom of expression within the law, and
>
> (b) protecting users from unwarranted infringements of privacy, when deciding on, and implementing, safety policies and procedures.

There seems to be a degree of wishful thinking in the claim that this and related provisions in the Bill would adequately protect freedom of expression. In particular, service providers are required only to 'have regard' to the right to freedom of expression and protecting privacy ('having regard' is the least onerous form of regulatory obligation), while having an absolute obligation to take steps to protect safety. If there is any arguable conflict between the two, there can be no doubt that the former will have to yield to the latter. In any case, there is no explicit duty within the Bill to require Ofcom to similarly protect freedom of speech and privacy when undertaking enforcement action.

The Bill's provisions on freedom of expression and on privacy constitute an admission that the drafters do not know how to remedy the risks that their own scheme is creating. Instead, the authors of the Bill pass the poisoned chalice to the service providers. The service providers will, in turn, provide convenient targets for criticism in case of inevitable scandals, both due to censoring and failing to censor online content.

To be able to show that they are acting 'proportionately' to remove illegal content and mitigate harmful content, the service providers will likely increase the use of automated tools to monitor user content. They will likely lean on the side of censorship to avoid the risk of large fines. Automated tools have well-known deficiencies (see, for example, Bloch-Wehba 2020; Douek 2021; Mchangama 2021; Shenkman et al. 2021):

- despite improvements, they still have issues with accuracy, and it is difficult to compare and benchmark accuracy of different tools;
- they cannot identify context ('An ISIS video looks the same, whether used in recruiting or in news reporting' (Keller 2018: 7));
- they 'can amplify social bias reflected in language' (Duarte and Llansó 2017: 4).

Hence, perfectly unobjectionable content is at risk of being at least flagged for manual review, if not simply removed or shadow banned. Human reviewers are not all-knowing, and they will inevitably make incorrect decisions about what material should be removed. And even if a human reviewer decides to restore content days or weeks later, this may be useless (e.g. for a user who attempted to participate in a real-time debate). Moreover, manual review of user content intended to be shared privately will surely infringe the user's privacy.

Technology warning notices and general monitoring of content

It is also likely that automated general monitoring of content will become an explicit regulatory requirement for at least some providers under Ofcom's codes of practice and enforcement of the Bill.

In their opinion on new Indian online safety legislation, UN Special Rapporteurs have expressed concern about

> a general monitoring obligation that will lead to the monitoring and filtering of user-generated content at the point of upload. This form of restriction would enable the blocking of content without any form of due process even before it is published, reversing the well-established presumption that States, not individuals, bear the burden of justifying restrictions on freedom of expression …[7]

Given Ofcom's responsibilities regarding terrorist and Child Sexual Abuse and Exploitation (CSEA) content, it is likely that the Bill's 'technology warning notice' procedure will be used by Ofcom to impose on service providers duties to engage in general monitoring of all user-generated content on their platforms or going through their messaging services. The fact that the automated tools for detecting terrorist or CSEA content are not good enough to avoid flagging mostly irrelevant content is unlikely to stop this.

Ofcom will be given a power to issue a 'technology warning notice' to a provider (clauses 63–65 of the Bill) if Ofcom believes that the provider's service has prevalent and persistent terrorist or CSEA content. Given the scale of use of the largest online platforms or messaging services, it is

7 Report of UN Special Rapporteur on the promotion and protection of freedom of opinion and expression, 11 June 2021 (https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=26385).

possible that Ofcom will interpret 'prevalence' broadly and decide that even though illegal content is a minuscule *proportion* of content on a given service, the *absolute* numbers are unacceptably high and that is sufficient for a regulatory intervention. Having decided that, Ofcom will be able to require a provider to 'use accredited technology to identify' terrorist or CSEA content and to take it down (clause 64). This is all the more likely given that Ofcom has already complained that it believes the threshold of 'prevalence' is too high, suggesting a strong institutional bias favouring 'safety' over privacy and freedom of expression.[8]

In practice, the use of 'accredited technology' will most likely mean not only that *all* video and photographic content transmitted through a service will be scanned by a machine – requiring content not to be end-to-end encrypted – but also that 'you can expect minimum-wage people in the Philippines to be viewing your naked kids' in innocently shared pictures with family members, as Alec Muffett, a former Facebook safety engineer, noted.[9]

Threat to end-to-end encryption

End-to-end encryption does immense good in protecting users. It is 'a basic and essential security protocol'.[10] It is as basic as using a good lock on one's front door. By analogy, just because there are children at home doesn't mean that the flat should not be allowed to have walls or locked doors. Just like with encryption, having walls makes it more difficult – but not impossible – for law enforcement to conduct investigations.[11]

8 Technical briefing note for the Joint Committee on the draft Online Safety Bill. *Ofcom*, 6 October 2021 (https://www.ofcom.org.uk/__data/assets/pdf_file/0037/226999/technical-briefing-joint-committee-online-safety-bill.pdf).

9 https://twitter.com/AlecMuffett/status/1406319821587947520

10 MPs: Encryption keeps your constituents safe, *Open Rights Group*, 14 June 2021 (https://www.openrightsgroup.org/blog/mps-encryption-keeps-your-constituents-safe/). See also: A 'key' for encryption, even for good reasons, weakens security, *New York Times*, 15 July 2016 (https://www.nytimes.com/roomfordebate/2016/02/23/has-encryption-gone-too-far/a-key-for-encryption-even-for-good-reasons-weakens-security).

11 Indeed, in the Schrems litigation the Court of Justice of the European Union has twice held ([2015] EUECJCJEU C-362/14 and [2020] EUECJCJEU C-311/18) the data protection law of the European Union is inadequate because it allows data transfers to the US, where it can be intercepted (without appropriate safeguards) by the US authorities. Encryption is the only meaningful means to address the concern identified by the Court. To restrict encryption leaves citizens entirely vulnerable to the privacy intrusions identified in these cases.

It is extremely worrying how poorly this seems to be understood among law enforcement and in policy circles.[12]

Law-enforcement officials often claim that the benefits of encryption can be preserved while giving law enforcement 'back door' access. This is false. Giving a kind of universal decrypting key to law enforcement or to the provider of an online service *always* creates a risk of abuse by criminals, agents of hostile nations or rogue employees with access.

The Government's guidance for online service providers states that:

> End-to-end encryption makes it more difficult for you to identify illegal and harmful content occurring on private channels. You should consider the risks this might pose to your users.[13]

Furthermore, Home Secretary Priti Patel has claimed that Facebook implementing end-to-end encryption for Messenger would put child safety in 'jeopardy'.[14] Law-enforcement agencies have, for many decades, been concerned about what is being said on platforms that they cannot easily monitor.[15]

The Bill does not expressly prohibit end-to-end encryption. But given the Government's attitude, likely to be shared by Ofcom, the Bill will probably be interpreted as severely discouraging if not effectively prohibiting end-to-end encryption, for example, on private messaging platforms like Facebook Messenger and WhatsApp. As we discuss below, age verification is unlikely to stop motivated children from using services. Hence, if the Government sees encrypted messaging as a 'higher risk feature'[16] that should not be accessible to children, all users may be deprived of it in the name of protecting children. Furthermore, the premise is that children

12 Fiona Hamilton, Facebook 'putting profit before welfare of children', *The Times*, 30 June 2021 (https://www.thetimes.co.uk/article/facebook-profit-before-welfare-children-paedophiles-abuse-82ncnwzrq).

13 Private and public channels: improve the safety of your online platform, *DCMS*, 29 June 2021 (https://www.gov.uk/guidance/private-and-public-channels-improve-the-safety-of-your-online-platform#harms-that-can-happen-on-private-channels).

14 Priti Patel: Facebook encryption plan 'must not hamper child protection', *BBC News*, 19 April 2021 (https://www.bbc.co.uk/news/technology-56795852).

15 Priti Patel v Facebook is the latest in a 30-year fight over encryption, *The Guardian*, 19 April 2021 (https://www.theguardian.com/technology/2021/apr/19/priti-patel-v-facebook-is-the-latest-in-a-30-year-fight-over-encryption).

16 Written evidence submitted by the Department for Digital, Culture, Media and Sport and the Home Office (OSB0011) to the Joint Committee on Draft Online Safety Bill's Legislative scrutiny inquiry (https://committees.parliament.uk/writtenevidence/38883/html/ at [18]).

should not have access to encrypted conversations and therefore children should have less privacy and security than adults. It is clear that the current debate about child safety has tipped in the direction of universal surveillance as a solution, as opposed to, for example, more targeted police work.

Age verification

The Bill requires services to treat users as if they are a child by default. The only way to display content that is not just child friendly (such as YouTube Kids) will be to undertake robust age verification – meaning companies asking users to enter their driver's licences, passports or credit cards to ensure that they only access age-appropriate services.

This will create an extraordinary privacy risk, due to the increased gathering of data required by private companies. Beyond the general annoyance to users of having to constantly re-enter confidential information, the requirements could effectively mean the end of online anonymity.

This will have a particular concerning effect on minority groups. For example, it has been claimed that:

> Growing calls to end anonymity online also pose a danger. Anonymity allows LGBTQ+ people to share their experiences and sexuality while protecting their privacy and many non-binary and transgender people do not hold a form of acceptable ID and could be shut out of social media.[17]

Even the most invasive age-verification techniques will be highly unlikely to stop motivated under-18s from accessing services not directed to them. The move to introduce age verification is therefore likely to seriously threaten privacy and add substantial inconvenience while doing little to prevent motivated underage individuals from accessing content.

17 Online Safety Bill gives legal basis for censorship of LGBT people, Stephen Fry and campaigners warn, *iNews*, 1 September 2021 (https://inews.co.uk/news/online-safety-bill-would-give-legal-basis-for-censorship-of-lgbt-people-stephen-fry-and-campaigners-warn-1178176).

Innovation

The Bill creates an extraordinary set of duties on companies of all sizes, particularly with respect to risk assessments, as well as the expansive 'safety' mandate. Larger companies are likely to have the resources to develop policies and procedures, hire moderators and develop artificial intelligence to comply with the law. However, having to repeat and update risk assessments with every technological change will be a serious obstacle to innovation, especially by smaller (the most dynamic) companies.

The Government's impact assessment indicates that the proposals will cost £2.1 billion, with an extraordinary £1.7 billion expected to be spent on content moderation. Even then, this is likely to be a substantial underestimate of the regulatory costs on innovation, competition and smaller companies. These costs will be crushing for start-ups and scale-ups, cementing the power of Big Tech.

Facebook founder Mark Zuckerberg warned a US congressional inquiry that:

> When you add more rules that companies need to follow, that's something that larger companies like ours just have the resources to go do and it just might be harder for a smaller company just getting started to comply with.

Facebook is a multi-billion-dollar company that can afford to comply with government regulation in numerous countries by hiring thousands of censors. It is the smaller, newer companies that will struggle to moderate potentially offensive material.

It will be nigh on impossible for smaller firms to fully comply with this legislation, particularly start-ups who are entering the market with limited resources. This could ultimately lead to a substantial decrease in the willingness of investors to enter the online space in the UK, seriously undermining the broader goals of the Government to promote competition in digital markets. Start-up trade body the Coalition for a Digital Economy (Coadec) found that 68 per cent of UK investors would respond by reducing investment in local platform businesses because of increased liability.

The Government claims that the 'the Bill takes a proportionate and risk-based approach'.[18] However, this means that the Bill simply contains vague provisions that businesses need to take 'proportionate' measures considering, among other things, 'the size and capacity of the provider of a service'. This is not a clear, operationalisable standard that a small business or a start-up could easily apply. Instead, businesses will require costly specialist legal advice to assess what measures are 'proportionate'. And they will need this advice not once, but every time they introduce any operational or technological changes that could potentially affect the assessment.

The businesses that will benefit from the Bill are law firms and, as the Government likes to point out, firms from the 'the safety technology sector'.[19] The problem with seeing the latter as a genuine benefit for the UK economy is that at least some of the products offered by those firms continue to have little market demand in the absence of a law forcing their adoption. And it is highly debatable whether, for example, age-verification services, which will inevitably create new privacy and security risks, provide social benefits to outweigh those risks.

There is also a substantial risk that many foreign companies respond to the regulatory risk presented by the Bill by choosing to not operate in the UK. This can be operationalised by 'geoblocking' access to British users – as was the case for many American sites following the introduction of the GDPR. Alternatively, the regulator may choose to block these sites. As well as undermining the free flow of information to the UK, this would limit the ability of British users to access newer and smaller platforms, adversely affecting competition within the industry.

18 Written evidence submitted by the Department for Digital, Culture, Media and Sport and the Home Office (OSB0011**)** to the Joint Committee on Draft Online Safety Bill's Legislative scrutiny inquiry (https://committees.parliament.uk/writtenevidence/38883/html/ at [12]).

19 Written evidence submitted by the Department for Digital, Culture, Media and Sport and the Home Office (OSB0011) to the Joint Committee on Draft Online Safety Bill's Legislative scrutiny inquiry (https://committees.parliament.uk/writtenevidence/38883/html/ at [13]).

Independent Reviewer of Online Safety Legislation

There are therefore serious concerns that, even if one assumes the best intentions of the Government and future regulator (Ofcom), the online safety regime could have significant negative consequences for freedom of expression, innovation and the quality of digital services available in the UK.

The authors of the draft Bill opted for a relatively high level of abstraction, leaving key details to be determined after the Bill becomes law. Those details will determine whether the potential negative consequences just mentioned will be realised. To a large extent it will be for the regulator, Ofcom, to make those key decisions. However, Ofcom will not have any incentive to treat freedom of speech, protecting privacy, and promoting a vibrant digital economy as seriously as the aim to ensure online safety. Given that those goals are in tension, it is difficult to expect one organisation – such as Ofcom – to retain an even-handed attitude to the competing aims instead of associating its mission more closely with one goal (promoting online safety).

A recent speech by Ofcom Chief Executive Melanie Dawes about safety and regulation did not mention freedom of expression.[20] This could indicate which values the regulator will prioritise. Moreover, high-profile incidents may put Ofcom under political pressure to take a tough line. For example, in the aftermath of the death of Sir David Amess, there were attempts (apparently advanced on no factual basis) to link the murder with social

20 Digital is not a sector – why regulators must collaborate for a safer life online, *Ofcom*, 6 October 2021 (https://www.ofcom.org.uk/about-ofcom/latest/media/speeches/2021/collaborating-for-safer-life-online).

media activity. Such cases seize the public and political attention whereas the routine abrogation of freedom of expression may cause little outcry, even if its total effect is egregious. Ofcom will inevitably be more swayed by the former than the latter.

Thus, if the potentially damaging regulatory tools in the Bill are to become law, there is a need to create an independent, permanent and adequately resourced mechanism of oversight, to help ensure that the crucial interests likely to be given less weight by the regulator (freedom of speech, privacy, innovation) are safeguarded.

The Independent Reviewer of Terrorism Legislation

This is a similar role to that undertaken by Independent Reviewer of Terrorism Legislation. As described by the current officeholder:

> The Independent Reviewer's role is to inform the public and political debate on anti-terrorism law in the United Kingdom. I do this in the regular reports that are prepared for the Home Secretary or Treasury and then laid before Parliament, in evidence to parliamentary committees, in articles and speeches, in media interviews and debates, in posts on this website and via twitter (@terrorwatchdog).[21]

The Reviewer undertakes an annual review of the operation of various pieces of terrorism legislation, publishes one-off reports instigated by the Reviewer or ministers, provides evidence to Parliament, and writes articles and speeches. It is a three-year Public Appointment, provided with administrative assistance and a Special Adviser. The Reviewer does not provide redress for individual incidents, but rather, takes a broad analytical approach.

Applying this model to online safety would raise less complex issues than in the terrorism case since, for the most part, it would not require access to secret and sensitive national security information.

21 The Independent Reviewer's role, *Independent Reviewer of Terrorism Legislation* (https://terrorismlegislationreviewer.independent.gov.uk/about-me/).

An Independent Reviewer of Online Safety Legislation and Regulation

An Independent Reviewer of Online Safety Legislation and Regulation could play an important role, as set out below.

Purpose: to ensure that the online safety regime promotes freedom of expression and privacy, innovation and the quality of digital services available in the UK.

Powers

The Reviewer would have powers, with respect to the online safety regime, to:

- monitor Ofcom's enforcement activity, including interviewing Ofcom employees and accessing any internal Ofcom documents the Reviewer considers relevant;
- scrutinise the Secretary of State's activities in respect to digital regulation, including instructions provided to Ofcom;
- mandate that Ofcom and the Secretary of State respond to requests for information from the Reviewer and to respond to their recommendations;
- advise, formally and informally, Ofcom and the Secretary of State;
- scrutinise and recommend changes to proposed codes of conduct and secondary legislation before presentation to Parliament;
- communicate findings to Ofcom, the Secretary of State, Parliament and directly to the public;
- gather views and data from civil society organisations, academia and industry on a voluntary basis;
- instigate investigatory proceedings against Ofcom both in respect of policy issues and individual decisions and also intervene in appeals brought by service providers or users in respect of Ofcom decisions;
- undertake any further activities necessary to fulfil the purpose of the Reviewer.

Staff and resourcing

The Reviewer should be supported by an adequately staffed and resourced office, independent both from Ofcom and from the Government. The Reviewer would need to have both expertise and a clear mandate to promote freedom of speech, privacy and innovation, and thus be well-placed to engage formally and informally with Ofcom to advise on issues within the Reviewer's remit. Having such an independent advisor would help Ofcom to better realise the difficult mission that involves conflicting goals.

However, limiting the Reviewer's role to advising Ofcom would be insufficient for creating a desired incentive structure within the online safety regime. By communicating directly with Parliament and the public, the Reviewer would be able to inform public debate, which could have beneficial effects on the practice of the online safety regime.

Knowing that actions that hinder privacy, innovation or freedom of expression will be publicly discussed by the Reviewer would create an incentive, otherwise absent, for Ofcom to give more weight to those issues and potentially to refrain from some such actions. Moreover, by publicising the negative effects of the online safety regime on issues within the Reviewer's remit, the Reviewer would contribute to continuous post-legislative scrutiny of the future Online Safety Act, which may inform potential changes to the Act. Just like the terrorism legislation reviewer, it could lay reports before Parliament and engage with policymakers in an ongoing manner.

The Joint Committee on the draft Online Safety Bill proposed resolving oversight issues by establishing a standing parliamentary scrutiny committee.[22] However, this seems likely to prove inadequate for several reasons. A parliamentary committee, made up of MPs, risks falling prey to current political pressures and panics as well as partisan tendencies. The committee, made up of MPs with a plethora of other responsibilities, would also likely lack the time or resources to adequately scrutinise the regulator's actions. It would not narrowly focus, or perhaps not focus at all, on issues such as freedom of speech, privacy or innovation that are currently under-protected in the Bill.

22 No longer the land of the lawless: Joint Committee reports, *Joint Committee on the draft Online Safety Bill*, 14 December 2021 (https://committees.parliament.uk/committee/534/draft-online-safety-bill-joint-committee/news/159784/no-longer-the-land-of-the-lawless-joint-committee-reports/).

The Reviewer would be particularly important considering the ongoing political pressure to 'strengthen' the legislation with respect to safety.[23] If the Bill is made into law, the Internet will still not be a 'safe' place. This could lead to demands for even harsher legislation and less balancing of freedoms, privacy and innovation. An independent reviewer could be one of the few formal bulwarks against demands for more infringements on these paramount values, mitigating the risk of overstepping by the Secretary of State and Ofcom.

23 See, for example, Online Safety Bill: Committees warn Prime Minister over lack of action on harmful paid-for scam adverts, *Joint Committee on the draft Online Safety Bill,* 23 July 2021 (https://committees.parliament.uk/committee/158/treasury-committee/news/156885/online-safety-bill-committees-warn-prime-minister-over-lack-of-action-on-harmful-paidfor-scam-adverts/). Online Safety Bill: Culture Secretary Nadine Dorries vows to get tough on tech firms – as executives could face jail for breaches, *Sky News*, 4 November 2021 (https://news.sky.com/story/online-safety-bill-culture-secretary-nadine-dorries-vows-to-get-tough-on-tech-firms-as-executives-could-face-jail-for-breaches-12459767) and PM urged to enact 'David's law' against social media abuse after Amess's death, *The Guardian*, 18 October 2021 (https://www.theguardian.com/uk-news/2021/oct/18/pm-urged-to-enact-davids-law-against-social-media-abuse-after-amesss-death).

Other safeguards

In addition to improving oversight, there are a number of additional mechanisms that policymakers could take to improve the underlying Bill:

- Remove 'legal but harmful' content from the scope of content included in the Online Safety Bill, and focus instead on unlawful content. Unlawful content should have a higher threshold, such as replacing 'reasonable grounds' with 'clear illegality manifest on the face of the content'.
- If legal speech is not removed from the Bill's scope, the definition of 'harm' should be extremely limited and specific, clearly targeted, well-defined and set out in primary legislation.
- Mandate Ofcom to protect freedoms of speech and association as a paramount value when designing codes of practice with respect to online safety. Not interfering with the freedoms of speech and association of citizens should be the foremost responsibility of the regulator.
- Availability of an independent tribunal to appeal decisions made by Ofcom in respect of codes of practice, decisions and notices – instead of appeals only on the limited judicial review principles as the Bill now envisages. Affected service providers, users and the Reviewer should all be able to bring such appeals. Moreover, the Reviewer should be able to intervene in cases brought by service providers or users. In particular, Ofcom codes of practice should be challengeable where they adversely affect freedom of speech or privacy of users.
- Remove special provisions for journalists and democratic content. Instead, Ofcom should have an explicit duty to ensure their codes of conduct and enforcement do not infringe on legislative and common law rights to freedom of speech that protect all users and legal speech.

- Include additional parliamentary oversight with respect to categories of content and codes of practice.
- Remove private messaging entirely from the scope of the legislation, do not require age verification and do not place requirements to scan encrypted messaging services. Explicitly state that no part of the Bill intends to impose an obligation on a provider for general monitoring of user content.
- Limit the territorial scope to services established in the UK and providing services to UK residents: (1) service established in the UK or (2) positive conduct that targets a service to the UK. Make clear that, for example, the presence of subject matter of interest to people in the UK is not sufficient to amount to targeting; nor should targeting be inferred from the fact that a service is not geo-fenced.

Further research is required on these issues, which should be given greater consideration by the government in light of the recommendations of the Joint Scrutiny Committee.

Conclusions

The need for an oversight body independent from Ofcom stems from the conflict at the heart of the Online Safety Bill. On the one hand, there is a demand for safety, on the other a (much weaker) demand to protect freedom of expression and privacy, all while ensuring continuing innovation. It is unreasonable to expect a single organisation to address these conflicting goals, particularly without sufficient independent input. Given the importance of these questions – freedom of speech, privacy, innovation – an independent reviewer could be a proportionate response that does not contradict or diminish the goals of the draft Bill. In fact, it could strengthen the goals by ensuring they are properly balanced.

If some of the dangerous provisions in the Bill become law, the independent reviewer could contribute to building the public case for the law to be changed. However, it is also possible that the reviewer will influence Ofcom for the better by advocating internally for the values in the reviewer's remit and, if needed, by bringing into daylight the deficiencies of Ofcom's internal processes.

References

Bloch-Wehba, H. (2020) Automation in moderation. *Cornell International Law Journal* 53: 41–96.

Chomanski, B. (2021) The missing ingredient in the case for regulating Big Tech. *Minds and Machines* 31(2): 257–75.

Douek, E. (2021) Governing online speech: from 'posts-as-trumps' to proportionality and probability. *Columbia Law Review* 121(3): 759–834.

Duarte, N. and Llansó, E. (2017) Mixed messages? The limits of automated social media content analysis. Center for Democracy & Technology.

Keller, D. (2018) Internet platforms: observations on speech, danger, and money. Hoover Institution Essay.

Keller, D. (2021) *Amplification and Its Discontents*. Knight First Amendment Institute at Columbia University.

Llansó, E., Van Hoboken, J., Leerssen, P. and Harambam, J. (2020) Artificial intelligence, content moderation, and freedom of expression. Working Paper, the Transatlantic Working Group on Content Moderation Online and Freedom of Expression (https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf).

Mchangama, J., Alkiviadou, N. and Mendiratta, R. (2021) Rushing to judgment: are short mandatory takedown limits for online hate speech compatible with the freedom of expression? Copenhagen: Justitia.

Shenkman, C., Thakur, D. and Llansó, E. (2021) Do you see what I see? Capabilities and limits of automated multimedia content analysis. Center for Democracy & Technology.

Tylecote, R., Hewson, V., Lesh, M. and Harris, B. (2021) You're on mute: the Online Safety Bill and what the Government should do instead. Free Speech Union.

The Institute of Economic Affairs
2 Lord North Street
London SW1P 3LB
Tel 020 7799 8900
email iea@iea.org.uk