

Written evidence submitted by the British and Irish Law, Education and Technology Association (BILETA) (OSB0073)

Prepared on behalf of the British Irish Law, Education and Technology Association (BILETA) by Dr Kim Barker, Dr Guido Noto La Diega, Dr Ruth Flaherty & Dr Aysem Diker Vanberg.

The British and Irish Law Education Technology Association (BILETA) was formed in April 1986 to promote, develop and communicate high-quality research and knowledge on technology law and policy to organisations, governments, professionals, students and the public. BILETA also promotes the use of and research into technology at all stages of education. The present inquiry raises significant questions relating to the proposed Online Safety Regime, the intended regulatory body, and the scope of content within the measures proposed by the Bill. The present call for evidence raises technological and legal challenges that our membership explores in their research. As such, we believe that our contribution will add significant value to the scrutiny of the Draft Online Safety Bill.

Summary

- (i) We agree that there is a need to address the regulation of online speech, and online content. We also accept, and are supportive of the need to enhance protection of vulnerable users online.
- (ii) That said, we have some serious concerns over the proposed Draft Online Safety Bill (OSB), both in terms of its substantive aim, but also it's likely practical implications.
- (iii) The policy intention is clear, but it is not similarly clear as to how it is intended that this legislation will operate in practice.
- (iv) The Draft OSB lacks clarity in relation to how it will operate in relation to some elements of free speech, especially in the areas of democratic and journalistic content.
- (v) It is our view that the Draft OSB is a work in progress at best, and poses significant risks to content, but also expression rights. We also retain concerns as to the choice of regulatory body, and the likely enforceability requirements that will be needed.

We have directed our comments to: Objectives of the legislation; Content in scope; and other points of relevance.

Objectives

1. Will the proposed legislation effectively deliver the policy aim of making the UK the safest place to be online?

1.1. No. The Draft OSB Bill in its current form will not deliver the policy aim.

- 1.2. Whilst the intention behind the proposed legislation is ambitious, there are some serious concerns as to the efficacy of the Draft Online Safety Bill (OSB), and its aim of 'making the UK the safest place to be online'. This is first, due to its broad scope, and second, to its potential adverse implications on free speech. The Draft OSB will dramatically increase the amount of the red tape and bureaucratic burden for service providers, platforms and OFCOM.
- 1.3. Despite some alterations, the Draft OSB continues to define harm very vaguely particularly when it comes to legal but harmful content. Section 46(3) of the Draft Bill defines legal but harmful content as " content having, or indirectly having, a significant adverse physical or psychological impact on an adult of ordinary sensibilities." This is a definition that requires further elaboration and explanation to avoid misinterpretation and potential abuse.
- 1.4. The meaning of indirect harm also requires further explanation as this can be used by service providers and platforms to moderate and delete user content by applying a very low threshold.¹ This vagueness creates the danger of censoring a vast amount of content that is neither illegal nor harmful. In other words, in contrast to the suggestions of the Secretary of State², the Draft OSB risks being a censor's charter. Due to the hefty fines and other liabilities introduced in the Bill, platforms and service providers are likely to take down content without fully investigating whether it is harmful or not, overzealously to protect their own interests. This will have serious consequences for freedom of expression and on media plurality, encouraging the silencing of controversial and minority (opposition) opinions which are much needed in society.
- 1.5. The imposition of criminal liability on senior managers in the Bill is likely to have a chilling effect on free speech³. To tackle online abuse, there is a need for an ongoing and constructive dialogue between the Government and other stakeholders including senior managers working for service providers. Instead of holding the corporation to account, holding senior manager's liable for illegal and harmful content will have a detrimental effect on the development of this dialogue.
- 1.6. The categorisations of online abuse as outlined in the current draft legislation are also overly narrow, and fail to account for a broader range of harms experienced by those abused online.⁴ The conceptualisation of online safety, as given effect by the proposed

¹ Edina Harbinja, 'The UK's Online Safety Bill: Safe Harmful, Unworkable?' (18 May 2021) <<https://verfassungsblog.de/uk-osb/>> accessed 7 September 2021.

² Department for Digital, Culture, Media, Sport, 'Oliver Dowden's Opinion Piece for The Telegraph on the Online Safety Bill' (11 May 2021) <<https://www.gov.uk/government/speeches/oliver-dowdens-opinion-piece-for-the-telegraph-on-the-online-safety-bill.>> accessed 7 September 2021.

³ Open Rights Group, "Online abuse why management responsibility is not the answer (5 May 2021) <[https://www.openrightsgroup.org/blog/online-abuse-why-management-liability-isnt-the-answer/.](https://www.openrightsgroup.org/blog/online-abuse-why-management-liability-isnt-the-answer/)> accessed 7 September 2021.

⁴ Kim Barker & Olga Jurasz, 'Text-Based (Sexual) Abuse and Online Violence Against Women: Toward Law Reform?', Jane Bailey, Asher Flynn, and Nicola Henry (Eds.) *The Emerald International Handbook of Technology Facilitated Violence and Abuse (Emerald Studies In Digital Crime, Technology and Social Harms)*, Emerald Publishing Limited, (2021), pp. 247-264. <https://doi.org/10.1108/978-1-83982-848-520211017.>

legislative provisions is also, as a result, overly narrow with almost a singular focus and is unlikely to make the UK the safest places to be online for all – rather, it may – at best – become the safest place to be online for some.

- 1.7. The Draft OSB places a considerable burden on OFCOM, and service providers. OFCOM is required to publish various codes of conduct and it is required to enforce the Online Safety Bill, with its limited resources. On the other hand, service providers irrespective of their size (the proposed Bill does not specify any thresholds and applies to any company with insignificant size and revenue) are required to determine on a day-to-day basis what constitutes harmful content which will create an unprecedented amount of work and require resources. This may in the long run be detrimental to the innovation in digital services and may even harm the quality and diversity of service providers as they may choose to not operate in the UK.

2. Is the “duty of care” approach in the draft Bill effective?

- 2.1. Proposing a duty of care as the model for controlling speech online raises some significant concerns. The model of ‘duty of care’ is something that admittedly appears in other areas of law, such as that of occupiers liability, but that should not mean that it is a model that ought to be transferred to other areas. The duty of care proposal here is little more than a ‘pandora’s box’ of future problems.
- 2.2. The Duty of Care Model as proposed in the Draft OSB is, at best, an outline framework. There is little detail in the Draft OSB that indicates how the duty will operate. The overwhelming concern is that given the absence of detail, and the imposition of various duties (under the duty of care model) will almost certainly lead to censorship, over-removing, and as Smith comments, ‘collateral damage.’⁵ In the pursuit of online safety, by imposing a number of duties, the Draft OSB serves to pose threats to content that is legal, and not necessarily harmful. It leaves determinations of legality of content to private, and profit-driven entities.
- 2.3. The Index on Censorship has stated that the Draft OSB would be “catastrophic for ordinary people’s freedom of speech.”⁶ These concerns are also our concerns. The lack of specifics attached to the framework in its current form means that platforms will have little choice but to be proactive in determining which speech is to be deleted. This is likely to result in significant legal challenge, but also significant damage to the freedom of expression rights we all enjoy. In being asked to make determinations of legal speech, commercial platforms are being trusted with decisions on what is – or is not – permitted speech. The model proposed therefore rests on trust, placing the

⁵ Graham Smith, ‘Harm Version 3.0: the draft Online Safety Bill’ Cyberlegale Blog, 16 May 2021 <<https://www.cyberleagle.com/2021/05/harm-version-30-draft-online-safety-bill.html>> accessed 7 September 2021.

⁶ Index on Censorship, ‘Right to Type: How the “Duty of Care” model lacks evidence and will damage free speech’ 23 June 2021. <<https://www.indexoncensorship.org/2021/06/governments-online-safety-bill-will-be-catastrophic-for-ordinary-peoples-freedom-of-speech-says-david-davis-mp/>> accessed 7 September 2021

operators of platforms in a position where they are directly controlling the speech of an individual.⁷

2.4. The duty of care model proposed is therefore, far too idealistic. It is also something which poses greater risks than are acceptable in its current form. Simply because this model works in other situations, does not mean it is transplantable into the sphere of online speech.

3. Does the Bill deliver the intention to focus on systems and processes rather than content, and is this an effective approach in moderating content? What role do you see e.g., safety by design, algorithmic recommendations, minimum standards, default settings?

3.1. The proposed Bill mainly concentrates on content moderation. Hence it is difficult to say that the Bill delivers the intention to focus on systems and processes.

3.2. The Bill should place more emphasis on safety by design and on setting minimum standards rather than emphasising moderating content.

4. Does the proposed legislation represent a threat to freedom of expression, or are the protections for freedom of expression provided in the draft Bill sufficient?

4.1. The draft bill adopts a very broad and vague notion of harmful content which could have significant repercussions for online users⁸. While the draft legislation does provide a bit more detail than the White Paper, it will be supplemented by Codes of Practice to be developed by Ofcom, as well as secondary legislation which will further stipulate what constitutes harmful content. In this respect, the definitions in the Bill of harmful content, whether to children or adults, remain vague and unhelpful to guide service providers.

4.2. The draft Bill is likely to give unprecedented censoring powers to private companies which are neither desirable nor sustainable in the long run, especially given the flaws in the proposed duty of care model (see above at 2.1 – 2.4).

4.3. Section 9(3) of the Draft Bill imposes:

“A duty to operate a service using proportionate systems and processes designed to—

(a) minimise the presence of priority illegal content;

(b) minimise the length of time for which priority illegal content is present;

(c) minimise the dissemination of priority illegal content;

(d) where the provider is alerted by a person to the presence of any illegal content, or becomes aware of it in any other way, swiftly take down such content.”

⁷ Kim Barker, ‘Taming our Digital Overlords: Tackling Tech Through ‘Self-Regulation’? in Ignas Kolpkas and Julija Kalpokiene (eds) *Intelligent and Autonomous* (Brill, forthcoming 2022).

⁸ Christoph Schmon, ‘UK’s Draft Online Safety Bill Raises Serious Concerns Around Freedom of Expression’ (14 July 2021) <<https://www.eff.org/deeplinks/2021/07/uks-draft-online-safety-bill-raises-serious-concerns-around-freedom-expression#>> accessed 7 September 2021.

As pointed out by Smith, pre-Brexit, the duties under Section 9(3) (a) to (c) of the Bill would have been contrary to the ECommerce Directive's Article 15, as the above provisions impose general monitoring obligations on hosting providers.⁹

Content in Scope

5. The draft Bill specifically places a duty on providers to protect democratic content, and content of journalistic importance. What is your view of these measures and their likely effectiveness?

- 5.1. The Draft Bill is intended to create a framework for preventing online harms which is “coherent and comprehensive...proportionate, risk-based and tightly defined in its scope”¹⁰. Specific issues being targeted are keeping children safe from harmful material, preventing the spreading of hate messages and discriminatory communications, and protecting democracy¹¹.
- 5.2. We agree with other bodies such as the Electronic Frontier Foundation and New America’s Open Technology Institute that these are important aims¹² considering how much power online platforms now have in sharing communications and other information. However, these aims must be balanced against protecting freedom of speech and avoiding over-censorship.
- 5.3. The Draft OSB poses several key issues in relation to exempted content which, while aiming to protect freedom of speech, lead to concerns regarding the scope and application of the stated duty of care. As such, these concerns mean the censorship included in the Bill is currently disproportionate to the aims of the legislation.
- 5.4. The duty to protect democratic content is laid out in Section 13 of the Bill, which defines such content as either “news publisher content in relation to” the service provided, or “regulated content in relation to”¹³ that service, so long as it is also (or can be argued to be) “specifically intended to contribute to democratic political debate in the United Kingdom or a part or area of the United Kingdom”¹⁴. This

⁹ Graham Smith, ‘Harm Version 3.0: the draft Online Safety Bill ‘ (1 June 2021) <https://inform.org/2021/06/01/harm-version-3-0-the-draft-online-safety-bill-graham-smith/>.

¹⁰ Online Harms White Paper: Full Government Response to the Consultation, <<https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response>> accessed 7 September 2021, paragraph 17.

¹¹ Gov.uk Press Release, ‘Landmark laws to keep children safe, stop racial hate and protect democracy online published’ (12 May 2021), available at <<https://www.gov.uk/government/news/landmark-laws-to-keep-children-safe-stop-racial-hate-and-protect-democracy-online-published>> accessed 06/09/2021

¹² EFF and OTI Joint Comments in Response to UK Online Harms White Paper (2019) <https://newamericadotorg.s3.amazonaws.com/documents/UK_Online_Harms_White_Paper.pdf> accessed 7 September 2021.

¹³ S13(6)(a)(i)-(ii) of the Draft Bill

¹⁴ S13(6)(b) of the Draft Bill

definition would cover much online discussion on news sites, as well as other sites such as social media services and video sharing platforms.

- 5.5. The preliminary issue with this duty is that it is clearly only intended to apply to the largest service providers, referred to as Category 1 providers¹⁵. However, it is as yet unclear how OFCOM will define Category 1 providers, as the Bill only enables the Secretary of State to specify the conditions for each category, based on number of users and functionality¹⁶. While the accompanying press release makes clear that this is purely intended to be for “the largest and most popular social media sites”¹⁷, without further specific definition, these rules are too vague to be applied as they stand, in comparison to the more specific language used in the Digital Services Act.¹⁸
- 5.6. A definition for a Category 1 provider must be included in the legislation. This should ensure that sites can be reasonably certain which duties apply to them, preventing the otherwise attendant issues with placing heavy burdens on potentially miscategorised smaller platforms¹⁹. This is even more important given that the Bill does not focus on individual items of content, but is designed as an overarching framework. Without clarity in this area, we will have “a regulatory framework which will have the badge of a duty of care but will leave individuals distinctly confused about what that duty of care means for them”²⁰.
- 5.7. S12 is also of concern in relation to freedom of speech. The accompanying Press Release highlights the desire for the Bill to “ensure people in the UK can express themselves freely online and participate in pluralistic and robust debate”²¹. This is specifically stated in s13(3) to apply to a “diversity of political opinion” and is intended to cover such things as promotion of political parties or political issues ahead of democratic events or campaigning around issues. The duty is on Category 1 services not to remove this type of material. While this is an admirable aim, there are some notable concerns in relation to how this will interact with freedom of speech.
 - 5.7.1. The definition of ‘democratic content’ in s13 as ‘specifically intending to contribute to democratic political debate’ is again very broad. It is hard to see how this will

¹⁵ S5(5)(d) of the Draft Bill

¹⁶ Schedule 4(1)(1) of the Draft Bill

¹⁷ Gov.uk Press Release, ‘Landmark laws to keep children safe, stop racial hate and protect democracy online published’ (12 May 2021), available at <<https://www.gov.uk/government/news/landmark-laws-to-keep-children-safe-stop-racial-hate-and-protect-democracy-online-published>> accessed 7 September 2021.

¹⁸ While Brexit means the Act will not become law in the UK, the type of clear distinction it uses in relation to the services it covers¹⁸ is one which is to be recommended, given the stated desire of the Government for the Bill to be “risk-based” and “proportionate”¹⁸.

¹⁹ Christop Schmon, ‘UK’s Draft Online Safety Bill Raises Serious Concerns Around Freedom of Expression’ EFF.org, July 2021, available at <<https://www.eff.org/deeplinks/2021/07/uks-draft-online-safety-bill-raises-serious-concerns-around-freedom-expression>> accessed 7 September 2021.

²⁰ David Barker, ‘Online Harms – The good, the bad and the unclear’ Out-Law Analysis (18 February 2020), available at <<https://www.pinsentmasons.com/out-law/analysis/online-harms-good-bad-unclear>> accessed 7 September 2021.

²¹ Gov.uk Press Release, ‘Landmark laws to keep children safe, stop racial hate and protect democracy online published’ (12 May 2021), available at <<https://www.gov.uk/government/news/landmark-laws-to-keep-children-safe-stop-racial-hate-and-protect-democracy-online-published>> accessed 7 September 2021.

apply in relation to specific issues such as COVID-19, which is a highly contentious issue²², yet has also become a live political issue, and so it is unclear how it would be covered by this legislation.

5.7.2. It is also unclear how this will help prevent the dissemination of terroristic propaganda, as much of it is drawn up as legitimate political discourse²³.

5.7.3. The opposite concern is that not all democratic content can be directly tied to a single political debate, and it is hard to see how content that is arguably of democratic importance will be defined should it not link to a specific electoral issue.

5.8. Allied to these concerns is the continuing issue of content moderation. Policing of online material has been a struggle for many years²⁴, and issues have been raised regarding the proportionality of prosecution of online speech in relation to offline words²⁵. This has led to concerns that “lobby groups will be able to push social networks to take down content they view as not politically correct, even though the content is legal”²⁶. Without clear boundaries on how this definition applies to content, and where the content appears, there are likely to be similar issues surrounding the moderation of user-generated content under Article 17 of the Copyright in the Digital Single Market Directive, which has been highly contentious. As written, the Bill would similarly incentivise the use of filtering technologies, which are problematic at best under their current form²⁷. As such, a clearer statement on the boundaries of this form of speech, including a specific statement that the general duty of care will not require automated content filtering as part of the s7 Risk Assessment Duties. Failure to do so, means that important content will be removed, or refused to be hosted at all²⁸. The duty is on the service provider to prove content “is, or appears to be specifically intended”²⁹ to be of this type, which is a burden some service providers may not wish to take.

5.9. Journalistic content is also exempted from the duty of care for Category 1 providers under s14 of the Bill, which ensures these providers permit content either generated

²² and one which is subject to a vast amount of dis- and mis-information online

²³ Mohammedwesam Amer (2020), ‘Terrorism and social media discourse studies: Issues, challenges and opportunities’, in H. Wu Lee & M. Van de Logt (Eds.), ‘Liberal arts perspectives on globalism and transnationalism: Within the knot’ (1st ed, Cambridge Scholars Publishing, 2020) 148–166

²⁴ Kate Klonick, ‘The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression’ (June 30, 2020). Yale Law Journal, Vol. 129, No. 2418, 2020, Available at SSRN: <<https://ssrn.com/abstract=3639234>> accessed 7 September 2021.

²⁵ Paul Bernal, ‘Response to Online Harms White Paper’, (Paul’s Blog, July 2019), <<https://paulbernal.wordpress.com/2019/07/03/response-to-online-harms-white-paper/>> accessed 7 September 2021.

²⁶ David Davis MP, quoted in ‘Online Safety Bill ‘Catastrophic for free speech’ (BBC, June 2021) available at <<https://www.bbc.co.uk/news/technology-57569336>> accessed 06/09/2021

²⁷ Sabine Jacques, et al. ‘Automated anti-piracy systems as copyright enforcement mechanism: A need to consider cultural diversity’ European Intellectual Property Review 40(4) (2018); Ruth Flaherty, ‘Articles 11 and 13: Bad News for Some, or All of Us?’ (UEA Information Society Policy Blog, October 2018), available at <<https://ruthflaherty39.wordpress.com/articles-11-and-13-bad-news-for-some-or-all-of-us/>> accessed 7 September 2021.

²⁸ Patrick Maxwell, ‘The government’s online safety bill is another unseen power grab’ (Politics.co.uk, 9 Aug 2021), available at <<https://www.politics.co.uk/comment/2021/08/09/the-governments-online-safety-bill-is-another-unseen-powergrab/>> accessed 7 September 2021.

²⁹ S13(6(b) of the Draft Bill

by news publishers on their own site, or generated by users in response to that content³⁰. While it has been confirmed that citizen journalism will also be covered by this subsection., it is difficult to see how in practice content will be distinguished between journalistic and democratic, given how often the two overlap. For example, using the examples given above of COVID-19 and terroristic content, both seem to be permissible either as democratic content or as journalistic content, so long as the information contained within is framed as a piece of citizen journalism – i.e. as someone’s opinion on a newsworthy or political point. Given that the duty of care to make a dedicated and expedited complaints applies only to journalistic content, it is critical that this distinction is made clear.³¹ This complaints procedure should be available for both democratic and journalistic content, given the overlap between the two types of content (and the importance of both to freedom of speech).

5.10. The expedited complaints procedure in relation to journalistic content also raises further issues in relation to freedom of speech. Under s14(3) if content is taken down by a provider, they must make available to that user (or the content creator) a dedicated and expedited complaints procedure in relation to potential reinstatement. Yet, one of the fears already seen in relation to content moderation for user-generated content³² is that the sheer amount of online content means that there is no clear way to ensure this takes place swiftly, unless it relies upon automation instead of human interaction. The general duty to protect freedom of expression in the Bill is intended to potentially require human moderation³³, but it is not clear in the Bill how this will happen in practice.

5.11. A clear distinction must be made in the Bill between journalistic content and democratic content, along with reference to how the complaints procedure will work in practice.

6. Are the definitions in the draft Bill suitable for service providers to accurately identify and reduce the presence of legal but harmful content, whilst preserving the presence of legitimate content?

6.1. Section 39(2) defines ‘regulated content’, in relation to user-to-user services, as follows: ‘user-generated content, except— (a) emails, (b) SMS messages, (c) MMS messages (...).’ This means that, according to the current draft, the most popular private messaging services would be within the scope of the definition. In-scope private messaging would include standalone services such as WhatsApp, Telegram, and other messaging apps, or messaging utilities contained within other services, such

³⁰ S14(8) of the Draft Bill

³¹ S14(4) of the Draft Bill.

³² under Article 17 of the CDSM Directive

³³ Gov.uk Press Release, ‘Landmark laws to keep children safe, stop racial hate and protect democracy online published’ (12 May 2021), <<https://www.gov.uk/government/news/landmark-laws-to-keep-children-safe-stop-racial-hate-and-protect-democracy-online-published>> accessed 7 September 2021.

as Messenger, the private message function on Instagram, and other direct messaging on social media applications.

- 6.2. We agree with the Open Rights Group that ‘it is likely that service providers will be required to scan the contents of your private messages for subjective harms’,³⁴ and share concerns that this will mean the only way service providers can comply with the proposed legislation is for them to actively and repeatedly break end-to-end encryption and the protections it offers. We consider encryption as a fundamental guarantee for the human rights to privacy and freedom of expression. We therefore recommend that all private messaging be excluded from the scope of the proposed legislation and secondary instruments. Encryption should be legally incentivised, not hindered, to the benefit of individual rights and collective freedoms.
- 6.3. Section 41(2)(3) defines ‘illegal content’ as ‘content consisting of certain words, images, speech or sounds amounts to a relevant offence if the provider of the service has reasonable grounds to believe that (...) the use of the words, images, speech or sounds amounts to a relevant offence.’ There are mainly two issues with this definition.
 - 6.3.1. First, it is impossible to foresee its impact on fundamental rights, including freedom of expression and privacy, as the draft OSB leaves it to future secondary instruments to be adopted by the Secretary of State what is a ‘relevant offence.’ This flaw is not isolated as, most worryingly, it will be up to the Secretary of State to decide what is ‘priority illegal content.’ Under the Online Safety Bill services providers have to minimise the presence and the dissemination of priority illegal content (Section 9(3)). Although we do not know what ‘priority illegal content’ means, we already know that there is a high risk that, for providers to minimise its presence and dissemination,]generalised monitoring systems will be put in place, to the detriment of fundamental rights, including freedom of expression and privacy. This would be in direct contrast with retained EU law. Indeed, under the eCommerce Regulations, which transposed the eCommerce Directive in the UK, there is a ban on general monitoring obligations. The Court of Justice recognised that to impose these filters – as the UK would be de facto doing with the draft OSB – run counter not only free speech and privacy, but also free enterprise as smaller providers will not be able to afford compliance.³⁵ Alongside the threats to individuals and businesses, we urge the UK Government to reconsider these pre-emptive positive obligations on online providers as the divergence with the EU would constitute a hurdle in the cross-border trade with the main commercial partner of the UK.³⁶

³⁴ Heather Burns, ‘Encryption in the Online Safety Bill’ (*ORG*, 20 July 2021)

<<https://www.openrightsgroup.org/blog/encryption-in-the-online-safety-bill/>> accessed 7 September 2021.

³⁵ See *Scarlet Extended SA v SABAM* [2011] ECR I-11959; Guido Noto La Diega, ‘Grinding privacy in the Internet of Bodies. An empirical qualitative research on dating mobile applications for men who have sex with men’, in Ronald Leenes, Rosamunde van Brakel, Serge Gutwirth and Paul De Hert (eds), *Data Protection and Privacy: The Internet of Bodies* (Hart 2018) 21-69.

³⁶ ‘Section 9(3) thus departs from 20 years of EU and UK policy aimed at protecting the freedom of expression and privacy of online users’ (Graham Smith, ‘Harm Version 3.0: the draft Online Safety Bill’ (*Cyberleagle*, 16 May 2021) <<https://www.cyberleagle.com/2021/05/harm-version-30-draft-online-safety-bill.html>> accessed 6 September 2021).

6.3.2. The second issue with the definition is that it outsources to private companies the decision on what is illegal. It is unclear what is the standard that providers have to meet to show that they had or had not reasonable grounds to believe that some multimedia content instantiated a relevant offence. It is foreseeable that – both for economic reasons and to present an appearance of objectivity of the moderation process – providers will put in place AI or otherwise automated systems to decide what is to be allowed to remain online.³⁷ These automated private policing activities are dangerous as research by the Center for Democracy & Technology convincingly shows:³⁸

- State-of-the-art automated analysis tools that perform well in controlled settings struggle to analyse previously unseen types of multimedia;
- Decisions based on automated content analysis risk amplifying biases present in the real world;
- Automated tools perform poorly when tasked with decisions requiring appreciation of context;
- Generalised claims of accuracy do not represent the actual multitude of metrics for model performance;

6.4. It is difficult to explain and account for the steps automated tools take in reaching conclusions. The risk of being handed high fines is likely to push providers to adopt restrictive measures and over-remove content in their efforts to meet their new online safety duties. The use of automated models will only increase these problems of over-removal and ultimately censorship. The Government seems to think that freedom-of-expression impact assessments can remove the problem. However, it is fair to say that it ‘smacks of faith in the existence of a tech magic wand (to posit that an) obligation to conduct a freedom of expression risk assessment could *remove* the risk of collateral damage by over-removal.’³⁹

6.5. When it comes to ‘illegal content’, our recommendations are:

- (i) not to leave its definition to secondary instruments;
- (ii) (ii) not to leave its definition to private companies;
- (iii) (iii) to rule out or at least disincentivise automated moderation and filtering;
- (iv) (iv) not to repeal the ban on general monitoring obligations;
- (v) (v) to exclude pre-emptive measures, including monitoring and filtering illegal content or at least to limits its operation to terrorist and CSEA.

6.6. These risks are particularly high for another category of content that is within the scope of the draft bill: legal but harmful content. This refers to content whose nature is such that ‘the provider of the service has reasonable grounds to believe that (...) there is a material risk of the content having, or indirectly having, a significant adverse

³⁷ To some extent the UK Government itself is aware of the risk but it does little if anything to minimise it. See DCMS, Impact Assessment RPC-DCMS-4347(2) [166].

³⁸ Shenkman, C., Thakur, D., Llansó, E. (2021) Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis. Center for Democracy & Technology.

³⁹ Smith (n5).

physical or psychological impact on an adult of ordinary sensibilities.⁴⁰ It should be said from the outset that we believe that the new legislation should apply only to illegal content and leave out the nebulous concept of ‘lawful but awful.’⁴¹ Either the content is actually harmful – in which case legislation should outlaw it – or it should remain within the domain of free speech: the UK Government should take responsibility in deciding what is admissible and what is not, it should not outsource the decision to the assumptions and evaluations of private companies that do not have the resources, skills, and more importantly the legitimacy to carry them out.

6.7. Unlike the current draft, in the Full Consultation Response in December 2020 the UK Government had maintained that the proposed duty of care should relate only to defined kinds of harm; although still subjective, that response was positively aligned to the subject matter of comparable offline duties of care. We would agree that the current wider definition ‘enables more vague and subjective kinds of harm to be brought back into scope of service provider duties,’⁴² which is a threat to legitimate free speech. These dangers of vagueness and subjectivity can be observed with particular clarity through the lens of the ‘person of ordinary sensibilities’, a legal fiction that would aim to increase the objectivity of the process of moderation of legal yet harmful content. It fails to achieve this aim, however.⁴³

6.7.1. First, the concept is taken from the abuse of private information, but in that context it operates in an entirely different manner as the person with ordinary sensibilities is the person whose information is disclosed not the person who receives the information.⁴⁴

6.7.2. Second, in importing this fiction the UK Government forgot to include a reasonableness requirement that is pivotal to the misuse of private information. If the psychological reaction to online content is unreasonable, the content should not be policed.

6.7.3. Third, the UK Government has been clear that psychological harm should not be limited to medically recognised conditions and should include significant negative effects on the mental state of an individual. It is hard to predict how this will work in practice but there is the risk that upsetting yet legitimate speech could be filtered out.

6.7.4. Fourth, what is worse, legal but harmful content includes content that has a mere indirect impact on the individual.⁴⁵ As an explanation of indirectly harmful content,

⁴⁰ Online Safety Bill, s 46(3).

⁴¹ Index on Censorship, ‘Government’s Online Safety Bill will be “catastrophic for ordinary people’s freedom of speech”’ (*Index on Censorship*, 23 June 2021) <<https://www.indexoncensorship.org/2021/06/governments-online-safety-bill-will-be-catastrophic-for-ordinary-peoples-freedom-of-speech-says-david-davis-mp/>> accessed 6 September 2021.

⁴² Smith (n 5).

⁴³ For further details on the following points see Graham Smith, ‘On the trail of the Person of Ordinary Sensibilities’ (*Cyberleagle*, 28 June 2021) <<https://www.cyberleagle.com/2021/06/on-trail-of-person-of-ordinary.html>> accessed 7 September 2021.

⁴⁴ *Campbell v Mirror Group Newspapers Ltd* [2004] UKHL 22.

the bill refers to ‘content causing an individual to do or say things to a targeted adult that would have a significant adverse physical or psychological impact on such an adult.’⁴⁶ This scenario rests on the untenable assumption that those who harm others after being exposed to content do so unwillingly.

- 6.7.5. Finally, the supposedly objectively standard of the person of ordinary sensibilities no longer plays a role when the provider ‘has knowledge, relevant to the content, about a particular adult at whom content is directed, or who is the subject of it;’⁴⁷ in that case, the provider can disregard the legal fiction and is expected to measure the harm with regard to the sensibilities of that particular person. There is the possibility that by simply notifying the provider about their specific characteristics and sensibilities, online users can effectively sidestep the ‘person of ordinary sensibilities’ standard thus increasing risks of censorship. Furthermore, the inclusion of ‘group characteristics’ in s45(4) poses additional risks, not least because there is potential for this to result in content being erroneously categorised as harmful.
- 6.8. Whilst we reiterate that the best option would be to remove altogether lawful content from the scope of the proposed legislation, as a second-best solution we recommend that (i) more objective standards are developed (ii) if the ‘person of ordinary sensibilities’ is retained, it should include a reasonableness requirement; (iii) indirect harms be excluded.

7. Final Remarks – The Suitability of OFCOM.

- 7.1. The choice of OFCOM as the likely regulatory body for the proposed regime is concerning. While it is a suggestion of convenience given that OFCOM is already established, it is not a suggestion that we have confidence in.
- 7.2. OFCOM is sorely lacking in staffing, and the requisite expertise to competently undertake the required regulatory role in the context of online safety. The recruitment and financial resources to allow this to be manageable are unlikely to come to fruition.
- 7.3. The regulatory body for media, should not be the same body charged with addressing online content regulation. Given the need for a bespoke, and nuanced approach to regulating online content, and especially online speech, OFCOM is not well suited to this.
- 7.4. Other options, including for example, the Information Commissioner’s Office (ICO) would be more suited to acting in a regulatory capacity here, given that it is an independent body with a history of balancing privacy needs with applying information and technology law online.

Individual supporters:

⁴⁵ Online Safety Bill, s 46(7)

⁴⁶ Online Safety Bill, s 46(7)(a).

⁴⁷ Online Safety Bill, s 46(6).

Dr Maureen O. Mapp, Birmingham Law School
Dr Michaela McDonald, Queen Mary's University, London
Dr Irene Couzigou, School of Law, University of Aberdeen
Dr Karen McCullagh, School of Law, University of East Anglia

September 2021