University of the Pacific Law Review

Volume 55
Issue 2 *Symposium Free Speech Forum*

Article 12

2-20-2024

# Regulation of Online Speech in the UK

Mariette W. Jones
*Middlesex University*

# Regulation of Online Speech in the UK

*Mariette W. Jones*[*]

TABLE OF CONTENTS



## I. INTRODUCTION

It is not a good time to voice a dissenting opinion in the United Kingdom, and especially not when that opinion is critical of certain powerful groups or individuals. In that case, you may lose your employment, find yourself uninvited or barred from events, or otherwise rendered an Unperson. Take the example of Professor Kathleen Stock, highly respected philosopher, academic, author, women's rights advocate, and recipient of the Order of the British Empire. In 2021, she was hounded out of her academic position after a prolonged, and sometimes physical, campaign of harassment by students, for daring to voice her opinion about gender self-identification.[1] Journalists are also in the firing line; Piers Morgan, a veteran presenter, was forced to resign after he expressed disbelief at a number of statements made by a member of the British Royal Family to an audience of millions during an interview by Oprah Winfrey. Love him or hate him, he did not do more than voice an opinion, and it must be scant comfort that afterward many statements in the Oprah interview were proven to be false.[2]

---

* Senior Lecturer in Law, Middlesex University, London.

[1] Richard Adams, *Sussex Professor Resigns After Transgender Rights Row*, GUARDIAN (Oct. 28, 2021), https://www.theguardian.com/world/2021/oct/28/sussex-professor-kathleen-stock-resigns-after-transgender-rights-row (on file with the *University of the Pacific Law Review*); *see* also Jane O'Grady, *What Does Kathleen Stock, Alleged 'Transphobe', Actually Believe?*, TELEGRAPH (Nov. 2, 2021), https://www.telegraph.co.uk/books/authors/does-kathleen-stock-alleged-transphobe-actually-believe/ (on file with the *University of the Pacific Law Review*).

[2] *See* Lucy Campbell, *Archbishop of Canterbury: Harry and Meghan's Legal Wedding Was on Saturday*, GUARDIAN (Mar. 30, 2021), https://www.theguardian.com/uk-news/2021/mar/30/archbishop-of-canterbury-harry-and-meghans-legal-wedding-was-on-saturday (on file with the *University of the Pacific Law Review*); Max Foster, *Critics Point Out Inconsistencies In Sussex's Oprah Interview*, CNN (Mar. 25, 2021), https://www.youtube.com/watch?v=PUMIYmlCJHQ (on file with the *University of the Pacific Law Review*) (although since hidden by CNN); Ben Hill, *CNN Accused of Censoring Its Own Report on Meghan Markle's Oprah 'Inconsistencies' As Critics Claim Network Was 'Got At',* SUN (May 4, 2021), https://www.thesun.co.uk/news/14851705/cnn-censor-report-meghan-markle-oprah-interview/ (on file with

British universities, the supposed bastions of liberal freedoms, frequently make the news for deplatforming speakers. The situation has become so lopsided that Parliament has been trying to force universities to allow a diversity of views on campuses, and has now come to the stage of contemplating legislation to mandate free speech on campus.[3]

Nor is the curtailment of speech a new issue in the UK. To take one example, free speech in the UK was chilled to such an extent by the common law tort of defamation that not only did the United States deem it necessary to enact legislation to protect their citizens' First Amendment rights against possible suit in the UK,[4] but the UK Parliament also enacted the UK Defamation Act of 2013, with the specific aim of redressing the balance between speech and reputation in favor of freedom of expression.[5] At the Second Reading of the Bill for this Act in the House of Commons the then Lord Chancellor and Secretary of State for Justice said:

> I share the mounting concern of recent years that our defamation laws … are at risk of damaging freedom of speech without affording proper protection. No one can be satisfied with a situation where the threat of lengthy and costly proceedings has sometimes been used to frustrate robust scientific and academic debate, to impede responsible investigative journalism and to undermine the good work undertaken by many non-governmental organisations.[6]

The very need for this major reform could, in hindsight, be seen as a significant red flag for free speech in the UK. It arguably does not go far enough to truly prioritize freedom of speech. As social media speech is conducted online, a good starting point is to examine this Act's relevant provisions, as these reflect the general "notice and takedown" regime in other UK statutes.

---

the *University of the Pacific Law Review*); *see also* Tom Bradby, *Prince Harry Denies He and Meghan Said the Royal Family Was Racist*, ITV (Jan. 8, 2023), https://www.itv.com/news/2023-01-08/prince-harry-denies-he-and-meghan-said-the-royal-family-was-racist (on file with the *University of the Pacific Law Review*) (In what can only be seen as an example of gaslighting, the main insinuation in the interview namely that the British Royal Family were racist, was later denied by Prince Harry himself.).

[3] Richard Adams, *Universities Minister: One Set of Guidelines on Free Speech Needed*, GUARDIAN (May 3, 2018), https://www.theguardian.com/education/2018/may/03/universities-minister-one-set-of-guidelines-free-speech-campus (on file with the *University of the Pacific Law Review*); JOINT COMMITTEE ON HUMAN RIGHTS, FREEDOM OF SPEECH AT UNIVERSITIES, 2017–19, HC 1279 & HL 162 (London); *Freedom of Expression: A Guide for Higher Education Providers and Students' Unions in England and Wales,* EQUAL. AND HUM. RTS. COMM'N (Feb. 2, 2019), https://www.equalityhumanrights.com/en/publication-download/freedom-expression-guide-higher-education-providers-and-students-unions-england (on file with the *University of the Pacific Law Review*).

[4] Securing the Protection of Our Enduring and Established Constitutional Heritage (SPEECH) Act of 2010, 28 U.S.C.A. § 4101 (passed by the Obama Administration after US academic author Rachel Ehrenfeld was successfully sued for libel in the UK on tenuous grounds (*Bin Mahfouz v Ehrenfeld* [2005] EWHC 1156 (U.K.)), making foreign libel judgments unenforceable in US courts, unless those judgments are compliant with the US constitutional protection of freedom of speech).

[5] MINISTRY OF JUSTICE, THE GOVERNMENT'S RESPONSE TO THE REPORT OF THE JOINT COMMITTEE ON THE DRAFT DEFAMATION BILL, 2012, Cm. 8295, at 4 (UK); Defamation Act 2013, c. 26, § 5 (UK), https://www.legislation.gov.uk/ukpga/2013/26/section/5.

[6] 12 June 2012, HC Deb (2012) col. 177 (UK).

The Defamation Act 2013 to a certain extent succeeds in its aim to lessen the chill to free speech,[7] but its provisions about potential online defamation miss the mark. The Act includes a new defence for operators of websites for actions that are brought in defamation in respect of a statement posted on the website. [8] The defense evolved from the common law defence of "innocent publication" and aims to protect those who do not have any editorial control over the material they handle. The defence extends to those who provide access to information on the internet where the information is provided by a person over whom the service provider has no control. In effect, the operator can raise as a defence that it was not the operator who posted the statement. If the real author cannot be identified (and therefore sued) by the claimant, however, the claimant will be entitled to complain to the operator and if the operator does not respond to the complaint, it may be sued in defamation. The important caveat here is that this defence is intended for the protection of facilitators of publication of defamatory statements not created by the operator itself. Malice defeats the defence, as would failure to respond to notices of complaint in accordance with the procedure set out in the *Defamation (Operators of Websites) Regulations 2013*.[9] The law around internet defamation is complex and evolving and this is reflected in the fact that these regulations, drafted to flesh out Section Five, are themselves complex and cumbersome.[10]

In practice this defence is no more than a 'notice and take-down" regime. It does not even change the common law very much. In *Byrne v Dean* [1937][11] the Court of Appeal held that the proprietors of a golf club who failed to remove an allegedly defamatory notice from the club's notice board had taken part in the publication of the notice by allowing it to rest on the wall, even though the notice had been placed on the wall without the consent of the proprietors. [12] The same reasoning applied to more recent internet cases such as *Tamiz v Google Inc* [2013][13] and so the result under the common law and the 2013 Act regime would for all intents and purposes be the same; the Internet Service Provider (ISP) could be held liable provided it had been placed on notice. One could argue although the defence is now formalised and stated more clearly, the result for free speech is unchanged or may actually be worse, as ISPs, instead of engaging with the merits of the notice, may just take the path of least resistance and remove all statements complained of without any investigation, or even worse, in an automated/algorithmic fashion.

---

[7] *See* Mariette Jones, *The Defamation Act 2013: A Free Speech Retrospective*, 24(3) 2 COMMC'NS L. 117–131 (2019) (UK).

[8] Defamation Act 2013, c. 26, § 5 (UK), https://www.legislation.gov.uk/ukpga/2013/26/section/5. This section reads as follows: (1) This section applies where an action for defamation is brought against the operator of a website in respect of a statement posted on the website. (2) It is a defence for the operator to show that it was not the operator who posted the statement on the website. (3) The defence is defeated if the claimant shows that— (a) it was not possible for the claimant to identify the person who posted the statement, (b) the claimant gave the operator a notice of complaint in relation to the statement, and (c) the operator failed to respond to the notice of complaint in accordance with any provision contained in regulations…

[9] The Defamation (Operators of Websites) Regulations 2013, SI 2013/3028, art. 3, (Eng. & Wales).

[10] *Id.*

[11] Byrne v. Dean [1937] 1 KB 818 (CA).

[12] It is likely that *Byrne v Dean* [1937] would be decided differently, if it were to come before the courts today, as section 10 of the 2013 Act would preclude jurisdiction to hear a defamation action against the proprietors of the club, other than where it was not reasonably practicable for an action to be brought against the author of the notice.

[13] Tamiz v. Google Inc. [2013] EWCA (Civ) 68 [44] (Eng. & Wales).

It seems as if free speech advocates in the UK are playing a long game of legislative Whack-a-Mole. No sooner have they achieved a victory in reforming a law to be less chilling of free speech that another pops up to restrict freedom of speech in another area: Draconian provisions and amendments in the areas of new anti-terror legislation,[14] rapidly expanding privacy and data protection[15] and the ever expanding regulation of "hate speech"[16] all have a direct impact on the ability to speak freely and to communicate ideas.

The UK is not alone in this arena, of course. Legal regulation of online content, under the guise of protecting citizens against "fake news" and a plethora of other perceived online threats, is being increasingly and fervently pursued across many, if not most, liberal jurisdictions.[17] These include recently enacted laws that amount to compulsory state-directed content moderation such as Germany's NetzDG ("Network Enforcement Act") 2017, [18] the EU's 2021 Regulation on online terrorism content and its Strengthened Code of Practice on Disinformation 2022, Australia's Online Safety Act 2021, as well as proposed laws such as the European Media Freedom Act which was proposed in September 2022. Australia seems to want to moderate content further than the provisions of that Act allows, as a bill focusing on misinformation specifically is currently being considered.[19] All of these focus on external regulation, i.e. states relying on privately-owned social media companies to control expression through the imposition of liability and a duty to actively monitor, control, and remove content.[20]

Perhaps in order to catch up with its European and other Western Liberal Democracy neighbors, the UK legislature set its sights well and truly on the online environment, and specifically social media platforms. The UK's recently enacted Online Safety Act of 2023 aims to go much further than the notice-and-take down regime that has made its insidious way into so much of European online content regulation.

---

[14] For an in-depth discussion of the far-reaching free speech implications of the UK's Counter-Terrorism and Border Security Act 2019 and its amendment of the Terrorism Act 200, see Eliza Bechtold & Gavin Phillipson, *Glorifying Censorship? Anti-Terror Law, Speech and Online Regulation*, THE OXFORD HANDBOOK OF FREEDOM OF SPEECH (Adrienne Stone & Frederick Schauer eds., 2021).

[15] On the proliferation of privacy and data protection suits, see Yanni Goutzamanis, *Closing the Floodgates on Privacy Class Actions: Lloyd v Google LLC*, 86(1) MOD. L. REV. 249–62 (2023) (UK).

[16] LAW COMMISSION, HATE CRIME LAWS, 2021, HC 942, at 1.35–1.37, 1.49, 4.220 (UK); *see* LAW COMMISSION, MODERNISING COMMUNICATIONS OFFENCES, 2021, HC 547 (UK).

[17] Gulizar Haciyakupoglu et al., *Countering Fake News: A Survey of Recent Global Initiatives*, RSIS, 22–27 (2018) (Sing.).

[18] Netzwerkdurchsetzungsgesetz [Network Enforcement Act], Sept. 1, 2017, FEDERAL LAW GAZETTE I at 3352, last amended by Act, July 21, 2022, FEDERAL LAW GAZETTE I at 1182, art. 3 (Ger.), translation at NetzDG - Act to Improve Law Enforcement in Social Networks, https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html (last visited Jan. 18, 2024; *see* Jacob Mchangama & Joelle Fiss, *The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship*, JUSTITIA 1 (2020) (Den.), https://justitia-int.org/wp-content/uploads/2020/09/Analyse_Cross-fertilizing-Online-Censorship-The-Global-Impact-of-Germanys-Network-Enforcement-Act-Part-two_Final-1.pdf.

[19] *See* Press Release, Michelle Rowland MP, New ACMA Powers to Combat Harmful Online Misinformation and Disinformation (Jan. 20, 2023), https://minister.infrastructure.gov.au/rowland/media-release/new-acma-powers-combat-harmful-online-misinformation-and-disinformation (on file with the *University of the Pacific Law Review*).

[20] Rep. of the Special Rapparteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, U.N. Doc. A/HRC/35/22, at 3 (2017).

## II. THE ONLINE SAFETY ACT

Almost a decade in the making, the Online Safety Act became law in September 2023.[21] Given what it aims to achieve, the slow progress of this behemoth was no surprise. Paul Wragg, for instance, had this comment on the aims of the Bill as initially set out by the UK Government:

> When we consult the white paper…we see a bewildering description of the problem. In the executive summary, the government describes it in five short paragraphs. But these paragraphs cover literally everything: gang culture, knife crime, terrorism, the manipulation of voters and destabilisation of democratic processes through fake news, child safety online, greater protection for online entrepreneurialism, consumer protection, paedophilia, self-harm, echo chambers, screen-time addiction, bullying, intimidation and harassment. Thus, it moves, awkwardly and confusingly, between criminality and immorality, between commerce and health and safety, between social cohesion and personal development. This is not a description of *a* problem, or even *some* problems. It is a description of *all* our problems.[22]

Four years and several rounds of consultations later, the finally promulgated Act is much amended—at least the draconian measures around "legal but harmful" online content have been deleted, albeit only as far as adults are concerned. However, the Act still suffers from over-reach and vagueness. Its core operational feature still entails the State compelling private companies to actively monitor, evaluate, and remove content created by third parties. The Online Safety Act establishes a new regulatory regime to address illegal and "harmful" content online, by means of imposing legal requirements about this kind of content on providers of internet user-to-user services and internet search engines. In order to enforce the newly created duties, the Act confers extensive new powers on the Office of Communications (OFCOM)[23] enabling them to act as the online safety regulator. This role includes overseeing and enforcing the new regulatory regime.[24]

Although the large multinational social media corporations such as Twitter, YouTube, Facebook, Instagram and so on are clearly targeted, Index on Censorship estimates that up to 180,000 big or small technology companies will fall under the remit of the Act. What is more, they will also be expected to implement it.[25] This means these ISPs will be mandated to restrict their users' freedom to both disseminate and access content online.

---

[21] Online Safety Act, (2023) (UK).

[22] Paul Wragg, *Tackling Online Harms: What Good Is Regulation?*, 2 COMMC'NS L. 49, 50 (2019) (Eng.).

[23] OFCOM, short for the Office of Communications, is a statutory regulatory body supervising the communications industry in the United Kingdom.

[24] Online Safety Act, (2023), Part 7 (UK). The Act also includes extensive provisions about regulating online pornography, which goes beyond the scope of this article.

[25] Index on Censorship, *Online Safety Bill Will "Significantly Curtail Freedom of Expression"* (May 4, 2022), https://www.indexoncensorship.org/2022/05/online-safety-bill-will-significantly-curtail-freedom-of-expression/ (on file with the *University of the Pacific Law Review*).

The UK is subject to the European Convention on Human Rights' Article Ten: Right to Freedom of Expression. In terms of Article Ten, Section One, the limited right is engaged when interfered with by a public authority or by a private body which is exercising public law functions.[26] Article Ten, Section T prescribes that free speech may only be interfered with if the interference is prescribed by law (the principle of legality), in pursuit of a legitimate aim, and necessary in a democratic society (the proportionality principle). As with all new legislation in the UK, section 19 of the Human Rights Act 1998 requires the Minister in charge of a Bill to make a statement as to whether it is compatible with Convention rights.[27] Despite Lord Parkinson of Whitley Bay duly making this statement on the final draft of the Online Safety Bill, opinions on the legality of several of the Act's provisions remain divided.

The regime enacted is not be the usual "notice and takedown" familiar from other legislation, but instead a proactive requirement to identify and remove certain content. The extent to which this is significant can be illustrated by comparing the underlying judicial starting points regarding notice and takedown regimes of the UK and the US, where the situation *vis-à-vis* ISPs differs greatly across the two jurisdictions. In the US, providers and users of interactive computer services, including ISPs, content hosts, and online search engines, enjoy complete statutory immunity from suit as regards the publication of information provided by others. For example, section 230(c) of the US Communications Decency Act 1996[28] contains a broad immunity from liability for internet intermediaries and others who facilitate or participate in the publication of online defamatory statements that they did not create. The section reads as follows, under the following heading:

> **Protection for "Good Samaritan" Blocking and Screening of Offensive Material**
>
> (1) Treatment of publisher or speaker
>
> No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.
>
> (2) Civil liability
>
> No provider or user of an interactive computer service shall be held liable on account of –
>
> (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
>
> (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).

---

[26] Wingrove v. The United Kingdom, App. No. 19/1995/525/611, ¶ 35 (Nov. 25, 1996), https://www.refworld.org/cases,ECHR,3ae6b6900.html (on file with the *University of the Pacific Law Review*).

[27] Human Rights Act, (1998) § 42, 19 (Eng.).

[28] Communications Decency Act, 47 U.S.C.A. § 230 (2018).

As can be seen, the section was intended to protect internet intermediaries who *removed* offensive material from their computer systems or websites from free speech suits by the providers of the material.[29] The starting point is therefore the assumption, stemming from and in accordance with the strong constitutional protection the First Amendment gives speech in the USA that speech should be allowed and not removed. Contrast that to the state in the UK requiring speech to be restricted on an *a priori* basis.

To do this, the Act imposes a duty of care analogous to the duty of care in tort or health and safety laws, but with a key difference that the duty would be based on a *precautionary principle*, i.e. by prior constraint.[30] Put another way, whereas in tort a wrongdoer is punished for breaking the law, here the balance is shifted to state-sanctioned private censorship of content, without any judicial consideration of whether the content was illegal. Index on Censorship rightly points out that this shift would mark "the most significant change in the role of the state over free speech in the UK since 1695."[31]

These duties of care are placed on ISPs in their capacity as private entities, but despite the language used in the Act, they mean that the ISPs will be required to act on behalf of the state to restrict freedom of speech.

The second far-reaching effect of the Act is that it largely exempts from its provisions content created by journalists and politicians.[32] In so doing, it creates two tiers of speech: free for journalists and politicians, and censorship for ordinary British citizens. This "journalistic content" exemption in the Act is unfair and unworkable. For a start, the traditional concept of a "journalist" is not exactly current anymore. Freelance writers, commentators, and bloggers may well argue that they also are journalists. Not only does this set a dangerous precedent by censoring citizen journalists and online news sources, [33] it also potentially engages issues surrounding access to cyberspace as a free speech issue. Access to cyberspace being a free speech issue and potentially growing in stature to a right, enjoys international recognition: The World Summit on the Information Society's (WSIS) "'Declaration of Principles' reaffirms Article 19 of the International Covenant on Civil and Political Rights (ICCPR) (right to freedom of opinion and expression) and emphasizes that communication is a basic human need that is central to the 'information society."[34]

The Act further places onerous duties on ISPs who are tasked with detailed and nuanced decisions about permissible speech, engaging issues that would be challenging even to senior judges. Surely nobody expects corporations to be as careful or informed as the judiciary when it comes to decisions about fundamental human rights. Companies exist to produce goods and to make profits in doing so. It is not fair to companies, their customers, or society as a whole to expect them to fulfil this role. The proponents of this

---

[29] Zeran v Am. Online, Inc., 958 F. Supp. 1124, 1134 (E.D. Va. 1997).

[30] Peter Coe, *The Draft Online Safety Bill and the Regulation of Hate Speech: Have We Opened Pandora's Box?*, 14(1) J. MEDIA L. 50 (2022) (UK).

[31] Written Evidence Submitted by Index on Censorship, INDEX ON CENSORSHIP 1 (UK Parliament, 2021), https://committees.parliament.uk/writtenevidence/41410/default/.

[32] *See, e.g.*, Online Safety Act 2023, §19, *Duties to Protect Journalistic Content*, and §17 *Duties to Protect Content of Democratic Importance*.

[33] *See* Index on Censorship, *supra* note 25.

[34] Dafna Dror-Shpoliansky & Yuval Shany, *It's the End of the (Offline) World as We Know It: From Human Rights to Digital Human Rights - A Proposed Typology,* 32(4) E.J.I.L. (2021), 1249–82, 1262 (2021) (Isr.).

legislation must recognize that outsourcing a judicial function to corporations is problematic. The government acknowledges this in a roundabout way. Consider this statement explaining one of the final amendments to the Online Safety Bill:

> New clause (NC14) establishes that providers' systems and processes should consider all reasonably-available contextual information when making judgements about whether content is content of a particular kind. This includes content judgements relating to other duties in the Bill (content of democratic importance, journalistic content, harmful-to-children and harmful-to-adults content), as well as content judgements relating to the illegal content duties.[35]

They end by stating, seemingly without irony: "This is important as it will often be difficult for providers to make judgements about content without considering context."[36] Perhaps they should have paused to reflect that even the police, and judges, find this kind of decision-making difficult.

The truth is, of course, that corporations are likely to either automate content moderation through the use of algorithms or alternatively, put some low-paid employee (or better still, an intern on no pay at all) to work on weeding out "illegal content." The digital arena the Act seeks to regulate is in fact already dominated by algorithms, automated decision making, and artificial intelligence in general to a very large extent. These all have certain characteristics that make them highly suited to the content moderation function mandated by the Act. Algorithmic decision-making is based on data gathering, processing, and analysis, which can predict human behaviour on the basis of scientific classifications and predictive formulas. Nevertheless, algorithms remain no more than "a list of instructions to be followed, like a recipe" and as such subject to the GIGO ("Garbage In, Garbage Out") principle.

An algorithm searches out and classifies data, an automated decision-making system "uses automated reasoning to aid or replace a decision-making process that would otherwise be performed by humans".[37] The transition from human to algorithmic decision-making also marks a shift from the exercise of public authority by public bodies to private entities, where decision making is opaque, hidden, usually not subject to appeal, and therefore blurs the lines of authority between government and technology companies.[38] It amounts to no less than a dehumanization of public authority.[39]

There are serious ethical questions about algorithmic decision-making, especially when the decisions affect subjects' human and other legal rights. It is an established principle of natural justice that the law should be certain and decision-making transparent. This implies the right not to be subject to an

---

[35] Online Safety Bill 2023, Fact Sheet on Changes to the Illegal Content Duties Within the Online Safety Bill (UK), https://www.gov.uk/government/publications/fact-sheet-on-changes-to-the-illegal-content-duties-within-the-online-safety-bill/fact-sheet-on-changes-to-the-illegal-content-duties-within-the-online-safety-bill (on file with the *Uiversity of the Pacific Law Review*).

[36] *Id*.

[37] AI NOW INSTITUTE, ALGORITHMIC ACCOUNTABILITY POLICY TOOLKIT 1–2 (2018), https://ainowinstitute.org/aap-toolkit.pdf (on file with the *University of the Pacific Law Review*).

[38] Helmut Phillip Aust, *Undermining Human Agency and Democratic Infrastructures? The Algorithmic Challenge to the Universal Declaration of Human Rights*, 112 AJIL UNBOUND 334 (2018).

[39] *See* Dror-Shpoliansky, *supra* note 34 at 1277.

automated judicial decision. However, algorithms are usually subject to proprietary software that are not always openly available for scrutiny, nor would they be comprehensible to most ordinary people if they were. This problem is known as the "algorithmic black box".[40] There are also questions about fairness and systemic bias—algorithms are after all only following a set of instructions coded by a human being.

Another key judicial safeguard to those subject to decision making is the right to know the identity of their judges. The use of algorithmic machines to assist or substitute human judicial decision-making therefore raises concerns about litigants' ability to access the reasons for the decision and to know who their judges are.[41] It also potentially falls foul of Section 49 of the Data Protection Act 2018,[42] which concerns the right not to be subject to automated decision-making, at least not if the decision is "significant",[43] meaning it either "produces an adverse legal effect concerning the data subject", or "significantly affects the data subject."[44] This somewhat circular and vague definition is open to interpretation.

The next major concern is that mass surveillance will in effect be mandated by the new law. New powers given to platforms mean restrictions on freedom of speech will be imposed by a private body exercising public law functions, and on a massive scale. In order to fulfil their onerous duties, ISPs will need to be able to actively monitor the content posted by their users. For instance, under the guise of "empowering users" in Section Fifteen, the Online Safety Act mandates services to make it possible for users to filter out content that encourages self-harm, suicide, eating disorders, or discriminates on a variety of grounds.[45] This begs the question of how the nature of this content will be ascertained in order for it to be filtered out. Surveillance is the only logical answer; Section 102 notices amount to state-mandated surveillance because they entail the right to impose technologies that would intercept and scan private communications on a mass scale. The Online Safety Act mandates state-backed surveillance of private communications which amount to some of the broadest and most powerful surveillance powers ever imposed in any Western democracy. They surpass even the provisions of the Investigatory Powers Act (2016), itself controversial enough to be colloquially referred to as the "Snooper's Charter."[46] Matthew Ryder KC, in his legal opinion on this aspect of the Online Safety Bill, rightfully pointed out that the very notion that the state can mandate the surveillance of millions of lawful users of private messaging apps should surely require a much higher threshold of legal

---

[40] Ric Simmons, *Big Data and Procedural Justice: Legitimizing Algorithms in the Criminal Justice System*, 15 OHIO ST. J. CRIM. L. 573, 573 (2018); *see also* Doaa Abu-Elyounes, *Contextual Fairness: Legal and Policy Analysis of Algorithmic Fairness*, 1 U. ILL. J.L. TECH. & POL'Y 1, 2–3 (2020).

[41] *See* Dror-Shpoliansky, *supra* note 34, at 1276.

[42] This Act is the UK counterpart of Article 22 of the EU General Data Protection Regulations. Article 22 provides data subjects with the right not to be subject to a decision based solely on automated processing, whenever such a decision "produces legal effects concerning him or her or similarly significantly affects him or her."

[43] Data Protection Act, (2018), § 49(1) (Eng.).

[44] Data Protection Act, (2018), § 49(2)(a) & (b) (Eng.).

[45] Online Safety Act, (2023), § 15 read with § 16(3)–(4) (UK).

[46] *See* Mariette Jones, *Double-Lock or Double-Bind? The Investigatory Powers Bill and Freedom of Expression In the United Kingdom*, *in* CYBERSURVEILLANCE IN A POST-SNOWDEN WORLD: BALANCING THE FIGHT AGAINST TERRORISM AGAINST FUNDAMENTAL RIGHTS (Russell L. Weaver et al. ed., 2017).

justification.[47] That such a raised threshold was not included in the enacted legislation should be of grave concern to free speech as well as privacy advocates. To put this in context, this level of state surveillance would only be possible under the Investigatory Powers Act 2016, for instance, if there is a threat to national security. The Investigatory Powers Act also at least contains some safeguards to protect the fundamental rights of UK citizens, something which the Online Safety Act lacks at a comparable level; this could leave it open to legal challenge, as it would run counter to, for instance, judicial oversight requirements established in UK common law.[48]

"User empowerment duties" placed on ISPs have further implications for the nature of online commentary. A feature of many social media platforms is that they allow anonymous users, or at least the use of pseudonyms. A central feature of the Online Safety Act, on the other hand, is that it mandates ISPs to "empower" users by making it possible for them to filter out "non-verified users."[49] So far so good as a means to filter content that a specific user does not want to see, or to counter trolling. But there is the possibility that this can be used by powerful people to prevent anonymous comment, often the only safe way to comment on these people.[50]

## III. DOES ONLINE SPEECH REALLY NEED THIS LEVEL OF REGULATION?

There are those who argue that although compulsory state-mandated content moderation, such as the statutes and bills discussed here, entail a form of direct (partial) state censorship, such formal legal schemes are justified because of mainly three reasons: (1) they are accompanied by legal remedies, (2) they engage the protections provided by national and international free speech guarantees and (3) they fall under the supervision of independent courts.[51] These three reasons are debatable, and also miss the point. It says that content moderation is justified because it can be *executed* in an (allegedly) controlled fashion. Given that one of the fundamental freedoms underpinning every liberal democracy is at stake, the only justification for censorship is an extremely important underlying reason. Is online speech really such a threat that the draconian measures outlined above are the only way forward? The fact that the law is workable does not answer the question whether it is desirable in the first place.

We are yet to see the *online nature* of communications to be the *cause* of many of the ills attributed to it. Consider the way in which the following opening paragraph of a typical article on the subject conflates correlation with causation:

> Developments in recent years have seen false information online evolve into a new phenomenon. What was once an issue that was

---

[47] MATTHEW RYDER KC, SURVEILLED AND EXPOSED: HOW THE ONLINE SAFETY BILL CREATES INSECURITY (Nov. 2022), Index on Censorship.

[48] Big Brother Watch v. United Kingdom, App. No. 58170/13, ¶ 450 (May 25, 2021).

[49] Online Safety Act 2023, (UK) § 15(10) read with §16(7) ("non-verified user" means a user who—(a) is an individual, whether in the United Kingdom or outside it, and
(b) has not verified their identity to the provider of a service).

[50] On the role and importance of anonymous speech to the right to freedom of expression, *see* ERIC BARENDT, ANONYMOUS SPEECH: LITERATURE, LAW AND POLITICS (Bloomsbury Publ'g 2016).

[51] *See* Eliza Bechtold & Gavin Phillipson, *supra* note 14, Ch 28.

> associated with trivial matters and minor harms, has transformed into something increasingly linked with serious matters and significant harms. This change can be attributed to two factors: a general increase in the use of social media communications for political strategy, and a more specific spike in the circulation of false cures and vaccine conspiracies from Covid-19. *The result of these developments, is a surge in associated physical, psychological, democratic and social harms*.[52]

We must ask serious questions around causation of concrete and quantified/quantifiable harm. This would be a logical counterfoil to the tort-analogous concepts such as a duty of care included in the Online Safety Act. On the same theme the nature of liability for online intermediaries, as proposed, closely resemble vicarious liability, which in tort is mainly strict liability. If we are moving in that direction further questions then arise about how to reconcile this with the duty of care, which is the starting-point for fault-based liability. It is submitted that without proper damage and causation elements, intermediate liability for online communications, as foreseen in the Online Safety Act, is far too draconian and will inevitably lead to a chill on free speech.

Some commentators argue that cyberspace is so unique that, when compared to current legal systems, there can be no question of normative equivalency. Cyberspace, according to them, needs its own set of human rights and special regulations.[53] Whether stricter regulation, with or without an accompanying set of "digital" human rights, would make people more or less likely to speak freely online is an open question. It could be argued that regulation coupled with the two-tier system of speech (journalists and politicians in the first-class carriage, the rest of the population in the third-class carriage) is only likely to further widen the digital divide. It is true that given the reach and universality of the internet, the scope for mischief is doubtlessly increased if not facilitated by the web. It may facilitate illegalities; it may amplify these. But the basic tenet of justice remains—fault must be attributed to the actual wrongdoer.

One must also further note the basic rule in terms of international law regarding online speech is that it is protected according to the same principles as offline speech.[54] The UN Human Rights Council, for instance, has long regarded any new rules which specially penalise online speech with great suspicion. Whilst it is recognised that in a few specialised areas such as sending spam emails, content restrictions may be needed; outside of these special cases, States should not create special Internet content restrictions.[55] As early as 2011, the United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States

---

[52] Henrietta Catley, *The Online Safety Bill: A Failure to Regulate False Information Online*, 28(1) COMMS. L. 23 (2023) (emphasis added).

[53] *See* Dror-Shpoliansky, *supra* note 34.

[54] Human Rights Council Res. 47/16, U.N. Doc. A/HRC/RES/47/16, at 3 (July 13, 2021).

[55] *Model Training Materials: Overview of Freedom of Expression Under International Law*, CTR. FOR L. AND DEMOCRACY, 17–18 (Oct. 2022), https://www.law-democracy.org/live/wp-content/uploads/2022/12/Training-Materials-1.FOE_.format-1.pdf (on file with the *University of the Pacific Law Review*).

(OAS) Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information issued a joint declaration on freedom of speech on the internet containing principles that run counter to those of the Online Safety Bill. In general, intermediaries, which provide merely technical services such as ISP, should be immune from any intermediary liability. Other types of intermediaries should also generally be shielded from liability.[56] Otherwise, the joint rapporteurs pointed out, private companies are likely to be overly zealous in restricting speech to protect themselves from liability, thereby inhibiting free expression. Any liability schemes should therefore not impose direct liability but only penalise a failure to act once properly notified about illegal content. They further pointed out that even this may create risks and accordingly, international standards are that intermediaries should never be required to monitor content proactively and should only be compelled to remove content upon receiving an order to do so from a court or other authoritative body.[57]

Noting concerns over harmful speech on platforms, the UN Special Rapporteur noted more recently:

> [T]he appeal of regulation is understandable. However, such rules involve risks to freedom of expression, putting significant pressure on companies such that they may remove lawful content in a broad effort to avoid liability. They also involve the delegation of regulatory functions to private actors that lack basic tools of accountability…. Complex questions of fact and law should generally be adjudicated by public institutions, not private actors whose current processes may be inconsistent with due process standards and whose motives are principally economic.[58]

Finally, while international standards call for careful review of any rules mandating content removal, they have increasingly embraced calls for regulating major tech companies in other areas. Specifically, they have called for greater transparency and procedural fairness and to protect human rights in the use of artificial intelligence.[59] Regulation in these areas can also raise complex challenges but it generally poses less of a risk to freedom of expression than content removal requirements. Further, transparency is crucial to understanding platform decisions, which restrict access to information online in ways that are relatively hidden, such as decisions to deprioritize certain content.[60]

---

[56] Frank LaRue et al., *Joint Declaration on Freedom of Expression and the Internet* (2011), http://www.law-democracy.org/wp-content/uploads/2010/07/11.06.Joint-Declaration.Internet.pdf (on file with the *University of the Pacific Law Review*).

[57] *Id.*, para. 2(b).

[58] Rep. of the S.R. on the Promotion and Protection of the Right to Freedom of Opinion and Expression, U.N. Doc A/HRC/38/35, at 17 (2018).

[59] For example, the civil-society led Santa Clara Principles, which are cautious of government regulation generally, suggest: "Governments and other state actors should consider how they can encourage appropriate and meaningful transparency by companies…including through regulatory and non-regulatory measures." *See The Santa Clara Principles on Transparency and Accountability in Content Moderation*, https://santaclaraprinciples.org/ (on file with the *University of the Pacific Law Review*).

[60] Centre for Law and Democracy, *supra* note 55, at 20.

## IV. REASONS TO PROTECT FREEDOM OF SPEECH

Discussion about online speech evoke novelty and the idea of uncharted territory. Idealistic lawmakers perhaps picture themselves as boldly going where no-one has gone before, rule-book in hand. Nevertheless, the principles underlying everything at stake here—the right to be tried by a jury of peers, the requirement that rules should be transparent, etc. dating back to Magna Carta (1215)—are ancient. The fundamental reasons why free speech should be protected are also very old, and bears repeating more often than of late. It is almost trite to state that freedom of expression is one of the most highly valued human rights, with almost universal acceptance as a *sine qua non* for democratic societies. It is included in international conventions such as the Universal Declaration of Human Rights (Article 19), the European Convention of Human Rights (Article 10) and the Charter of Fundamental Rights of the European Union (Article 11),[61] to name but a few. Free speech is protected in some form in every modern liberal democracy. It could even be argued that a state does not merit being called either liberal or a democracy, if it does not protect freedom of speech. Indeed, so well entrenched is this right that nowadays debate is usually constrained to its limitation, rather than its substantive nature. Because of its wide acceptance, Professor Barendt quite rightly remarks that free speech "is prized by liberals for reasons they may not understand."[62] Academic debates nowadays mostly seem to focus on reasons to curtail freedom of speech. Instead of freedom of speech, the de facto picture will be permitted speech. It is therefore vital that lawmakers should be reminded about the underlying reasons why freedom of speech should be highly valued and protected.[63]

The best place to start remains John Stuart Mill's classic essay *Of the Liberty of Thought and Discussion*, in his seminal treatise on democratic freedom, *On Liberty*.[64] Mill's stated goal in *On Liberty* is to identify the nature and limits of the power which can be legitimately exercised by society over the individual. His famous argument from truth emphasises the interests of society in discovering the truth. For this, he argues, society should not merely tolerate, but embrace speech that is considered objectionable. The reasons for this are fourfold. First, nobody is infallible, and therefore we must be open to the possibility that an opinion that deviates from the mainstream might be true. Next, even where an argument is substantially wrong, it may still contain a portion of truth that is missing from the accepted opinion. If the prevailing opinion is completely true, it still needs to be open to challenge for it is only through frequent challenge and vigorous defence that those who hold the opinion can fully understand the rational grounds for the opinion.

Finally, related to the last point and, of particular importance to our current analysis, Mills argues that in the absence of vigorous debate, the meaning of the doctrine itself will be in danger of being lost, or enfeebled, and deprived of its vital effect on character and conduct.[65] In short, an argument

---

[61] Article 11 states: "1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. 2. The freedom and pluralism of the media shall be respected."

[62] Eric Barendt, *Freedom of Speech* 1 (2005).

[63] *Id.* at Ch. 1.

[64] John Stuart Mill, *On Liberty* (1869).

[65] *Id*.

may persuade the audience members to change their minds, or it may cause them to defend their stance on the matter.

Then there is the argument of citizen participation in a democracy. Alexander Meiklejohn, a leading exponent of political speech, emphasized the importance of the electorate being able to access a variety of opinions on political and social matters. This at the least equals, and probably outweighs, the individual speaker's interest in participating in the discourse.[66] A further theory, best framed by Thomas Scanlon, explores free speech as an aspect of individual self-fulfillment or autonomy and holds that the justification for freedom of speech proceeds from the right of an individual to consider all the arguments and views that may determine their course of action.[67] Of course, the liberal notion of individual autonomy also includes the speaker's right to determine the content of their speech.[68] But, given the fact that the arguments of truth, citizen participation in democracy, and autonomy have been influential in shaping the development of constitutional free speech rights, it is fair to say that recipients, rather than speakers, are the primary object of free speech interests.[69] The categorisation of the right as a societal good is clear.

In various fora today, the argument that comes up most often is that in Mills' time social media platforms did not exist, and therefore neither the potential for harmful content to spread in the way and to the extent it can today. However, arguments about scale do not really affect the underlying principles.

There is also a troubling flavour of paternalism about the notion that today's citizens somehow need more protection from ideas floating about than did the citizenry in Mills' time. Access to the internet falls squarely under the protection of the right to freedom of expression expressed in Article 19 of the International Covenant on Civil and Political Rights because the internet is a medium that facilitates seeking, receiving, and imparting information and ideas. That is only realised if by "access" we mean full and equal access.

## V. CONCLUSION

Online speech in the UK is already being chilled due to the provisions and/or operation of a variety of laws. In statutes dealing with defamation, terrorism, data protection, privacy, and more, "notice and take-down" regimes and state mandated content moderation are to be found aplenty. This has reached such saturation that instead of thinking of speech as "free" one could be forgiven to observe that the adjectives that more accurately describe the situation in the UK nowadays is "restricted," "permitted," "proscribed," and "censored." What you can say now very much depends on who you are, what you are saying, and where you are saying it.

Against this background, the Online Safety Act goes even further. Instead of the usual notice and takedown regime, corporations are placed in charge of, legally required, to proactively seek out and limit online content. The likely result of outsourcing public duties like this would be censorship by algorithm. Various established principles of the law are being stretched or outright violated. The proposed law will not operate in a transparent or

---

[66] Alexander Meiklejohn, *Political Freedom* 64 (1960).

[67] Thomas Scanlon, *A Theory of Freedom of Expression*, 1 PHILOS. PUB. AFF. 204 (1972).

[68] *See* Ronald Dworkin, PHILOSOPHY OF LAW (1977) (Introduction).

[69] *See* Eric Barendt, *supra* note 50; Barendt, *supra* note 62.

accountable manner. It will be anti-democratic and it will violate the rule of law.

We are at the stage now where, as far as online speech in the UK is concerned, the default position is that of regulated—not free speech. It is time to fundamentally reconsider the direction in which the law is headed.

* * *