Is Free Speech Protected Under The Uk Online Safety Act 2023? Some Reflections On The Problem With Content Moderation On Social Media Platforms

Olufemi Ojo Ilesanmi[1], Chantal Mather[2]

[1]School of Law, Robert Gordon University, AB10 7QE, Aberdeen
[2]School of Law, Robert Gordon University, AB10 7QE,

Abstract
This study examines the problem with content moderation on social media platforms. In considering whether and to what extent free speech would be protected in the Online Safety Act 2023 of the United Kingdom, this article firstly explores free speech in general. Although free speech legislation and case law are relied upon in the arguments, this article focusses on the four commonly cited free speech arguments centring on: 1) truth, 2) self-fulfilment, 3) political participation, and 4) suspicion of government. It then attempts an investigation of the intent of Parliament in relation to the provisions of the law. Finally, it examines whether these provisions could protect free speech. The study finds that, although the law contains provisions relating to free speech, the attention it paid to cognate rights is uneven, such as the removal of illegal and harmful content, leaving 'free speech per se' as, seemingly, an afterthought.
Keywords: UK Online Safety Act 2023; Free Speech Theories; Content Moderation; Social Media; and Harmful Content

1. INTRODUCTION

The Online Safety Act 2023 received royal assent, becoming an Act in the UK law, on the 26th of October. This was preceded by a highly anticipated Bill,[1] which passed its final parliamentary debate on the 19th of September 2023. It was presented after scholars, such as Koltay[2] and Napoli,[3] expressed the view that Governments were not doing enough to combat illegal and harmful content online. The passed Bill, for example, contains provisions stating that illegal content[4] – that is content that amounts to a relevant offence

[1] For the 2023 Act, see An Act to make provision for and in connection with the regulation by OFCOM of certain internet services; for and in connection with communications offences; and for connected purposes. Available at https://www.legislation.gov.uk/ukpga/2023/50/enacted Accessed 03/11/2023. Hereafter described as 'the Act' or 'the Law.' It must be noted that the 2022 Bill, which was continually regulated, was introduced in the House of Lords on the 18th of January 2023 (HL Bill 87 (REV). Hereafter described as 'the Bill' or 'the Law' also. For its last version, see newbook.book (parliament.uk) Accessed 15/06/2023.
[2] Andras Koltay, 'Constitutional Protection of Lies?' (2020) Communications Law 25(3)
[3] Philip Napoli, 'What if More Speech is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble' (2018) Federal Communications Law Journal 70(55)
[4] Online Safety Act 2023, s 9. Last version Clauses 8(5)(g) and 9(2)(c).'Illegal because harmful' replaces 'legal but harmful'.

– should be removed by the internet service, as should harmful content,[5]which is described in the Bill as content that causes 'significant adverse physical or psychological impact' on a child or adult of 'ordinary sensibilities'.[6]

Since publishing the Bill in March 2022, many, such as Coe,[7] Trengove,[8] and Lesh and Hewson[9] argue that in its attempt to fight illegal and harmful content through the provisions, the Bill as passed compromises free speech. Free speech can be defined as the right to express any opinions without censorship or restraint,[10] and is generally regarded as one of the most important human rights, out of which all other human rights flow.[11] As it happens, when social media became established in the early 2000s,[12] the platforms were widely applauded for the positive effect they could bring to various human rights, including that of free speech.[13] In the words of Jack Balkin in 2004,

The digital age provides a technological infrastructure that greatly expands the possibilities for individual participation in the growth and spread of culture and thus greatly expands the possibilities for the realisation of a truly democratic culture[,][14]

showing that Balkin applauded social media for the possibilities of free speech, due to individuals now having the ability to participate more fully in democratic society through, for example, presenting their own views on their own accounts, instead of having to rely on traditional media, such as news outlets, to present their views.[15] That is not to say, however, that social media absolutely guarantees free speech,[16] instead it is important to remember various reasons as to why free speech might still be restricted, such as certain countries having restrictions on media freedom and therefore placing restrictions on what can be posted on social media.[17]

---

[5] Online Safety Act 2023, s 45-46.

[6] ibid. In the last version of the Bill, a search for 'significant adverse' returned only Clause 150 (c), which is about 'super-complaints'.

[7] Peter Coe, 'The Draft Online Safety Bill and the Regulation of Hate Speech: Have we Opened Pandora's Box?' (2022) Journal of Media Law 14(1)

[8] Markus Trengove et al, 'A Critical Review of the Online Safety Bill' (2022) Elsevier 3(8)

[9] Matthew Lesh and Victoria Hewson, 'An Unsafe Bill: How the Online Safety Bill Threatens Free Speech, Innovation and Privacy' (2022) IEA Briefing Paper

[10] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 2.

[11] Ammar Oozeer, 'Internet and Social Networks: Freedom of Expression in the Digital Age' (2014) Commonwealth Law Bulletin 40(2)

[12] Danah Boyd and Nicole Ellison, 'Social Network Sites: Definition, History, and Scholarship' (2007) Journal of Computer-Mediated Communication 13(1)

[13] Jack M Balkin, 'Digital Speech and Democratic Culture: a Theory of Freedom of Expression for the Information Society' (2004) New York University Law Review 79(1)

[14] Jack M Balkin, 'Digital Speech and Democratic Culture: a Theory of Freedom of Expression for the Information Society' (2004) New York University Law Review 79(1) 5.

[15] Jack M Balkin, 'Digital Speech and Democratic Culture: a Theory of Freedom of Expression for the Information Society' (2004) New York University Law Review 79(1)

[16] Kathleen Stock seems a plausible hypothesis that the intensity - and sometimes vituperative nature - of social media commentary inhibits free speech. For a recent discussion, see Learn to change your minds, Oxford University VC tells students (thetimes.co.uk) Accessed 15/06/2023

[17] Martin Scott, Mel Bunce, Mary Myers and Maria Carmen Fernandez, 'Whose Media Freedom is Being Defended? Norm Contestation in International Media Freedom Campaigns' (2023) Journal of Communication 73(2)

Nevertheless, living in a digital age of social media is paradoxical. While it allows for easier promotion of human rights and values in some cases,[18] media content can be harmful, or even illegal,[19] as mentioned above. This type of content is usually dealt with through content moderation, which can be defined as the 'screening, evaluation, categorization, approval or removal/hiding of online content according to relevant communications and publishing policies.'[20] This content moderation is usually done by the platforms themselves, through various techniques, such as algorithms and machine learning or simply through human moderators moderating pieces of content that have been flagged.[21] Nevertheless, there are arguments that social media platforms do not do enough to combat these online harms and that instead, governments should do more.[22] Hence the introduction of the Online Safety Bill, which, as shown in the Government White Paper,[23] aimed mainly at fighting online harms. However, just as scholars may argue that the provision is not doing enough,[24] some argue that in the attempt to combat and even prevent the abuse on online platforms, free speech is being compromised.[25]

This paper reflects on the problem with content moderation on social media platforms. In examining whether and to what extent free speech would be protected within the Law therefore, it firstly explores free speech in general. The paper then investigates the intent of Parliament in relation to the provisions of the 2022 Bill (as passed) and, by extension, the resultant 2023 Act. Finally, it examines whether these provisions do protect free speech. The study finds that, although the Law does contain provisions relating to free speech, the attention to other provisions is uneven, such as the removal of illegal and harmful content, leaving free speech as, seemingly, an afterthought.

---

[18] Jack M Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (2018) University of California David 51(1151)

[19] Andras Koltay, 'Constitutional Protection of Lies?' (2020) Communications Law 25(3); Philip Napoli, 'What if More Speech is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble' (2018) Federal Communications Law Journal 70(55)

[20] Terry Flew, Fiona Martin and Nicolas Suzor, 'Internet Regulation as Media Policy: Rethinking the Question of Digital Communication Platform Governance' (2019) Journal of Digital Media & Policy 10(1) 40.

[21] Kate Klonick, 'The New Governors: The People, Rules and Processes Governing Online Speech' (2018) Harvard Law Review 131 1598

[22] Philip Napoli, 'What if More Speech is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble' (2018) Federal Communications Law Journal 70(55)

[23] Harm to individuals is kind of immediate and focused. But erosion of free speech is relatively amorphous, but no less important. See Department of Digital, Culture, Media, and Sport, *Online Harms White Paper* (2020) <https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper> accessed 20 January 2023.

[24] Henrietta Catley, 'The Online Safety Bill: a Failure to Regulate False Information Online' (2023) Communications Law 28(1)

[25] Peter Coe, 'The Draft Online Safety Bill and the Regulation of Hate Speech: Have we Opened Pandora's Box?' (2022) Journal of Media Law 14(1); Markus Trengove et al, 'A Critical Review of the Online Safety Bill' (2022) Elsevier 3(8); Matthew Lesh and Victoria Hewson, 'An Unsafe Bill: How the Online Safety Bill Threatens Free Speech, Innovation and Privacy' (2022) IEA Briefing Paper

2. REVIEW OF RELEVANT LAWS AND THEORIES

Before it can be established whether, and to what extent, free speech is being protected within the Law, an overview of relevant concepts, theories and laws is considered. The Law itself is examined for an appreciation of its content. But first, the applicable laws and theoretical underpinnings for free speech discourse is briefly considered.

2.1 Free Speech

Free speech typically describes the right to express any opinions without censorship or restraint.[26] That is not to say, however, that free speech is an absolute right. Instead, it is often regarded as a qualified right, implying that certain limitations and restrictions are sometimes placed on what can be said.[27] For example, Article 10(1) of the European Convention on Human Rights (ECHR) gives a general right to freedom of expression,[28] with Article 10(2) subjecting this right to certain conditions, such as those that are prescribed by law (i), necessary in a democratic society (ii) or in the interest of (iii), for example, national security, public safety and for the protection of health and morals.[29] These three requirements are generally named the principle of legality (i), necessity (ii) and proportionality (iii).[30]

In the UK, the right to freedom of expression can be found primarily in the Human Rights Act 1998, where Schedule 1 provides for the same Articles found in the ECHR,[31] including Article 10 – that of freedom of expression.[32] This means that free speech is also a qualified right in the UK as it is 'subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society'. One might then ask, how these restrictions on speech play out in practical life. This has been an issue the courts have grappled with for years. In the older case of Handyside,[33] for example, it was held that even ideas and opinions that might offend, shock, or disturb, should still be protected under the principle of free speech, demonstrating that the court is reluctant in restricting speech even if it is for the protection of health and morals, one condition found in Article 10(2).[34]

Recent cases have held similarly. Scottow,[35] for example, concerned a case where Ms Scottow was convicted of an offence under the Communications Act 2003 for 17 social media messages, including messages about transgender issues, that Ms Hayden, a transgender woman, complaint had caused her annoyance, inconvenience, and needless anxiety. Nevertheless, on appeal the conviction was quashed with the judge stating that these types of discussions were important for political debate.[36] Similarly, in Miller[37] it was

[26] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 2.
[27] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 74.
[28] European Convention on Human Rights 1950, Article 10(1).
[29] European Convention on Human Rights 1950, Article 10(2).
[30] Rebecca Helm and Hitoshi Nasu, 'Regulatory Responses to 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation' (2021) Human Rights Law Review 21(2)
[31] Human Rights Act 1998, Schedule 1.
[32] Human Rights Act 1998, Schedule 1, Article 10.
[33] *Handyside v United Kingdom* [1976] 12 WLUK 53
[34] European Convention on Human Rights 1950, Article 10(2).
[35] *R v Scottow* [2020] EWHC 3421
[36] ibid.
[37] *R (on the application of Miller) v College of Policing* [2021] EWCA Civ 1926

held that the police arriving at the claimant's place of work, and warning the claimant about criminal prosecution because of his political opinions posted on social media, had a significant adverse effect on freedom of expression and that the political opinions stated in the case were contributing to an ongoing debate that was complex and multifaceted. Nevertheless, although Scottow[38] and Miller[39] show the court's reluctance to restrict free speech, the European Court of Human Rights recently implied in Delfi[40] that it is willing to limit the wide scope of free speech on the internet to protect other fundamental human rights, if the restrictions fall within Article 10(2). It will therefore be interesting to see how case law around internet services further develops, especially now with the introduction of the Online Safety Act.

Nevertheless, as stated in the introduction, the intention of this article is to focus mainly on the theory behind free speech, and how that might interact with the Online Safety Act. It is important therefore, to now consider the four commonly cited arguments in favour of free speech.

The first argument is often referred to as 'the argument from truth'[41] as a social good and is closely associated with John Stuart Mill. Mill argued that, if speech was restricted, discussion would not be possible and that it was the possibility of discussion that allowed for the truth to be discovered. Restriction on speech was therefore impermissible, because a restricted opinion might contain the truth.[42] Truth, according to Mill, is one of the fundamental ideals to be reached.[43] Of course this argument raises a few theoretical and philosophical questions First, what is truth?[44] After all, one person's truth can be the other person's falsity.[45] Secondly, is truth really the fundamental ideal to be reached?[46] Is it not so, that many societies want to protect other values as well? The most challenging ethical, philosophical, and legal problems arise when principles conflict. That throws the burden of presenting the truth, being a social good and virtue, onto decision-makers and encourages inquiries beyond consequentialism[47] and deontology, to virtue ethics. In the words of Max Weaver,[48]

---

[38] *R v Scottow* [2020] EWHC 3421

[39] R *(on the application of Miller) v College of Policing* [2021] EWCA Civ 1926

[40] *Delfi AS v Estonia* [2015] 6 WLUK 504

[41] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 7.

[42] John Stuart Mill, *On Liberty* (2nd edn, J W Parker and Son 1859)

[43] ibid.

[44] For a defence of the concept of truth, see Bernard Williams, *Truth and Truthfulness: an Essay in Genealogy* (Princeton University Press 2004)

[45] See, for example, studies on eyewitness testimony; Elizabeth Loftus and Jennifer Palmer, 'Reconstruction of Auto-Mobile Destruction: an Example of the Interaction between Language and Memory' (1974) Journal of Verbal Learning and Verbal Behaviour 13

[46] Henry John McCloskey, 'Liberty of Expression: its Grounds and Limits' (2008) An Interdisciplinary Journal of Philosophy 13(1)

[47] John Stuart Mill was a consequential list.

[48] Since questions arising from the truth arguments may not be explored in detail because of time and space, an overview of W D Ross' reflection on the value of truth as a social good might help. See generally Skelton, Anthony, "William David Ross", *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2022/entries/william-david-ross/>. Accessed 15/06/2023. One of the prima facie duties that Ross recognized was the duty of self-improvement, which he understood as the duty to increase one's own virtue or excellence.

Ross's view of virtue as self-control can be seen as a form of virtue ethics, which is a theory that focuses on the character and human flourishing of moral agents. Virtue ethics tells us that what matters most for ethics is not following moral rules or principles, but developing and applying the virtues that enable us to live well and contribute to the common good. Virtue ethics also emphasizes the role of emotions and friendship in the moral life, as well as the importance of practical wisdom for dealing with complex and context-sensitive ethical issues. Ross's virtue ethics can be contrasted with other forms of deontology or consequentialism, which tend to neglect or downplay these aspects of morality.
One only must think of racist speech being banned,[49] as well as certain advertising on drugs and tobacco,[50] to see what is meant here. Third, as Edwin Baker argues, 'why bet that truth will be the consistent or even the usual winner?'[51] Would it not be the case, that lies or falsities win the open and unrestricted discussion? Nevertheless, the idea that free speech leads to finding the truth is also found in case law, such as the case of Animal Defenders International,[52] where Lord Bingham noted that over time, in a public debate, 'the true will prevail over the false.'[53]

The second argument sees free speech as an essential aspect of individual's right to self-development and fulfilment.[54] When there are restrictions on speech, or even restrictions on what we are allowed to hear and read, it is argued that it inhibits our personality and growth.[55] However, this argument does not fully explain why free speech is so important to a person's self-fulfilment. For example, why are children shielded from some programmes on social media? Is it to shape their sense of themselves? After all, it is far from clear whether free speech always leads to personal happiness or that it satisfies any other basic human needs and wants.[56] Thomas Scanlon, responds to this in arguing that the autonomy of the person is of the utmost importance, and that a person can only be autonomous if he is free to weigh various arguments for various courses of actions against

---

For Ross, virtue was not only a means to an end, but also an intrinsic good that has value in itself. He defined virtue as "the disposition to act from the appropriate motives", such as the desire to do one's duty or to promote the good of others. Ross held that virtue has the highest value among all intrinsic goods, and that it belongs to a higher order of value than pleasure. He also suggested that virtue is closely related to knowledge, since both are forms of rational activity that express our nature as human beings.

[49] For example, the UK Public Order Act 1986, Part 3.

[50] For example, the UK Children and Families Act 2014, Part 5. Future studies might provide opportunities to expand the argument -e.g. free expression of racist ideologies having the potential to lead to violence or property damage, or creating a climate facilitative of unrest, internal repression and persecution, and social and economic reality, and so on.

[51] Edwin Baker, *Human Liberty and Freedom of Speech* (Oxford University Press 1989) 6. The theodicy here is even if truth wins in the end, untruth can do a huge amount of damage beforehand. He has a strong point.

[52] *R (On the Application of Animal Defenders International) v Secretary of State for Culture, Media and Sport* [2008] UKHL 15

[53] *R (On the Application of Animal Defenders International) v Secretary of State for Culture, Media and Sport* [2008] UKHL 15, para 28.

[54] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 13.

[55] ibid.

[56] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 14.

each other.[57] These arguments are only available if there is no restriction on speech. This argument also emerges in the UK case of Simms,[58] where the court observed that free speech had several objectives, including that of self-fulfilment.

The third argument is closely associated with Meiklejohn,[59] who said that the First Amendment[60] is there to protect the right of all citizens to understand political issues in order to participate in the workings of democracy.[61] This idea has also come back in case law, such as the famous United States case of Whitney v California,[62] in which Brandeis J stated '…that freedom to think as you will and to speak as you think are means indispensable to the discovery and spread of political truth.'[63] A few points are worth noting though. Just like the arguments above, this argument is in favour of discussion, and a wide range of views being available to the public. However, since this argument concerns citizen participation in a democracy, it only seems to cover political speech.[64] Thus, there would be little justification to extending the principle of free speech to that of, for example, literary and artistic use, or even pornography and commercial advertisement. Nor, and perhaps this is more important, would it protect speech that challenges the existence of a democratic government and its institutions,[65] or protect speech that is regulated or restricted by the elected representatives in a democracy?[66] Ronald Dworkin appears seems to disagree with that point of view, arguing that, in the terms of a constitutional democracy, political institutions must respect the rights of all citizens.[67] Everyone, including those concerning the minority, is entitled to participate in public discussion, as a result of which temporary political majorities are formed. This right, Dworkin argues, is so fundamental, that it cannot be surrendered to the powers of the elected majority.[68]

The last argument as primarily argued by Schauer is that one must be suspicious of the government,[69].[70] He points out that throughout history there have been attempts by powerful institutions, like the communist reign of Stalin,[71] as well as the Catholic Church,[72] to suppress speech. This raises the question of whether there is any speech that can be

[57] Thomas Scanlon, 'A Theory of Freedom of Expression' (1972) Philosophy and Public Affairs 204(1)

[58] *R v Secretary for the Home Department, Ex Parte Simms* [1999] UKHL 33

[59] Alexander Meiklejohn, *Free speech and its Relation to Self-Government* (Harper 1948); Alexander Meiklejohn, 'The First Amendment is an Absolute' (1961) Supreme Court Review 245

[60] Which is the Amendment that protects free speech in the United States.

[61] Alexander Meiklejohn, *Free speech and its Relation to Self-Government* (Harper 1948); Alexander Meiklejohn, 'The First Amendment is an Absolute' (1961) Supreme Court Review 245

[62] *Whitney v California* (1927) 274 US 357

[63] *Whitney v California* (1927) 274 US 357, para 42.

[64] Robert Bork, 'Neutral Principles and Some First Amendment Problems' (1971) Indiana Law Journal 47(1)

[65] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 19.

[66] ibid.

[67] Ronald Dworkin, *Taking Rights Seriously* (Duckworth Books 1977)

[68] Ronald Dworkin, *Taking Rights Seriously* (Duckworth Books 1977)

[69] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 21.

[70] Frederick Schauer, *Free Speech: A Philosophical Enquiry* (Cambridge University Press 1982)

[71] Michael Malice, *The White Pill: A Tale of Good and Evil* (Independently Published, 2022)

[72] Frederick Schauer, *Free Speech: A Philosophical Enquiry* (Cambridge University Press 1982)

restricted at all, for example, hate speech,[73]as explored by Jeremy Waldron. Furthermore, as Eric Barendt points out, 'is there a reason to be particularly suspicious of government regulation of free speech, compared with censorship or interference by other bodies such as churches, commercial companies, or even media corporations?'[74] In Europe, for example, there are cases[75] where the government has imposed on certain free speech rights regarding corporations such as the media, in order to protect the speech of others. Those cases are controversial in the US,[76] however, where the court is often reluctant to impose limits on the speech of, for example, the media.

Thus, as can be seen, there is a wide variety of material on free speech, whether that is through legislation or case law, or on the theory behind free speech. Before this article can move on, however, to determine whether, and to what extent, free speech is protected within the Online Safety Act, it is instructive to consider complementary or competing points of view in other writings.

2.2 The UK Online Safety Law

Peter Coe offers a significant critique of the passed Bill, outlining its advantages and disadvantages.[77] For example, in his article 'Misinformation, Disinformation, the Online Safety Bill and its Insidious Implications for Free Speech,'[78] Coe argues that in its attempt to combat misinformation and disinformation, the Bill does not have sufficient regard to freedom of expression, and that platforms are more likely to over-remove content in fear of the huge fines that can be imposed on them, than to pay sufficient importance to free speech with 'soft-duties' like 'having regard to' or 'taking into account' freedom of expression and the importance of journalistic content. Similarly, in his article 'The Draft Online Safety Bill and the Regulation of Hate Speech: Have we Opened Pandora's Box?'[79] Coe compares the Online Safety Bill with similar legislation in Germany and argues that the UK Bill, that is now law, does not pay enough importance to free speech with provisions such as 'taking into account free speech.' Although Coe's work is stimulating, its primary focus is on free speech in general, not the theory behind it. (This article seeks to address that omission.)

---

[73] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 21. Compare with Jeremy Waldron, 'Dignity and Defamation: The Visibility of Hate' (2010) Volume 123 Harvard Law Review 1596. Available at https://harvardlawreview.org/print/vol-123/dignity-and-defamation-the-visibility-of-hate/#:~:text=In%20his%20three%202009%20Holmes,of%20each%20member%20of%20society. Accessed 15/06/2023.

[74] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 22.

[75] Such as, but not limited to, *R (On the Application of Animal Defenders International) v Secretary of State for Culture, Media and Sport* [2008] UKHL 15; Case 131/12 Google Spain SL, Google Inc v Agencia Espaniola de Proteccion de Datos [2014] ECR 620; German case of 7 BVerfGe 198, 208; French case of Decision 84-181 of 10-11 Oct 1984, Rec 73.

[76] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 22.

[77] Peter Coe, 'Misinformation, Disinformation, the Online Safety Bill, and its Insidious Implications for Free Speech' (2021) Communications Law 26(3); Peter Coe, 'The Draft Online Safety Bill and the Regulation of Hate Speech: Have we Opened Pandora's Box?' (2022) Journal of Media Law 14(1)

[78] Peter Coe, 'Misinformation, Disinformation, the Online Safety Bill, and its Insidious Implications for Free Speech' (2021) Communications Law 26(3)

[79] Peter Coe, 'The Draft Online Safety Bill and the Regulation of Hate Speech: Have we Opened Pandora's Box?' (2022) Journal of Media Law 14(1)

Another scholar is Trengove,[80] who examines the Bill's rationale, its scope in terms of lawful and unlawful harms intended to be regulated, and how it will be enforced. In examining this, Trengove finds that further evidence is necessary to justify the Bill's interventions; that the Bill limits Parliamentary scrutiny and therefore risks democratic shortcoming; that the duties of the Bill may be too wide; and that the Code of Practice might be insufficient in terms of enforcement. Although Trengove argues that the Bill requires further refinement to protect free speech, Trengove does not explore the theory behind free speech.

Lesh and Hewson also offer their view on the Bill. In their paper 'An Unsafe Bill: How the Online Safety Bill Threatens Free Speech, Innovation and Privacy,'[81] it is argued that the Bill raises significant issues for free speech, and that there is lack of evidence as to how the provisions in the Bill will solve the Bill's objectives. When it comes to free speech specifically, the authors go into detail as to why the Bill is a risk to free speech, as too much content might be removed, or because it limits access to information. Nevertheless, free speech theory is scarcely mentioned.

For his part, Stefan Theil[82] specifically examines prominent free speech philosophers' points of view centring on the 'harm principle' – in other words, the principle that free speech may only be interfered with by the state if it meets a certain threshold of harm. Nevertheless, although Theil focuses on how the harm principle interacts with social media, it focuses mostly on how the private parties themselves, such as the social media platforms, might have too much power in deciding what is and is not allowed on their platform. Theil does not address the four commonly cited arguments justifying free speech specifically, nor does he focus on the Online Safety Bill.

3. THE ONLINE SAFETY LAW – WHY WAS IT INTRODUCED AND WHAT DOES IT CONTAIN?

As made apparent in the preceding section, there is useful scholarly commentary on the Online Safety Bill as passed, and its possible implications for free speech. Nevertheless, literature is lacking on how the Bill, that is now law, might interact with the four commonly cited arguments justifying free speech. Thus, in responding to whether, and to what extent, free speech is protected within the Online Safety Act, the following sections focuses on theory, although some legislation and case law are mentioned in passing. Before the question can fully be addressed, however, it is important to now see what the rationale behind the Act is, as well as the Act's intended operation practice.

3.1 Rationale

Until very recently, the problems associated with social media were managed through traditional laws, such as racial hate speech, which is criminalised under Part 3 of the Public Order Act 1986,[83] or defamatory statements on websites being required to be removed

[80] Markus Trengove et al, 'A Critical Review of the Online Safety Bill' (2022) Elsevier 3(8)

[81] Matthew Lesh and Victoria Hewson, 'An Unsafe Bill: How the Online Safety Bill Threatens Free Speech, Innovation and Privacy' (2022) IEA Briefing Paper

[82] Stefan Theil, 'Private Censorship and Structural Dominance: why Social Media Platforms should have Obligations to their Users under Freedom of Expression' (2022) Cambridge Law Journal 81(3)

[83] Public Order Act 1986, Part 3.

under section 5 of the Defamation Act.[84] These were supplemented by so-called 'soft-law'[85] in the form of self-regulation by the social media platforms themselves, through content moderation in various forms, such as algorithms or human content moderators.[86]
Nevertheless, because of the extent of online harms, as evidenced in the Government White Paper,[87] and the platforms, according to the Government, failing to adequately regulate these harms, the UK Government introduced a Bill designed to combat relevant harms.[88] In the White Paper, one can find the key objectives of the legislation, including:
A free, open and secure internet, freedom of expression online, and an online environment where companies take effective steps to keep their users safe, and where criminal, terrorist, and hostile foreign state activity is not left to contaminate the online space.[89]
In addition to that, the Government presents, amongst others, child sexual exploitation, the sale of opioids online, self-harm and suicide, online disinformation, cyberbullying, and online manipulation as the online harms that the Act is designed to address.[90]

3.2 Practice

How then, is the Act designed to combat these harms? In first instance, it is important to note that not all internet services fall under the scope of the Act. Instead, the Act identifies two categories: user-to-user services and search services.[91] User-to-user services are defined as 'internet services that allow users to generate, upload, or share content that can be encountered by other users on that service.'[92] Social media platforms would be an example of this. In addition to that, the Act also identifies Category 1 services, defined as 'high risk and high reach,'[93] and includes platforms such as Twitter, Facebook and Instagram.[94] Furthermore, for an internet service to fall under the scope of the Act, it needs to have a link with the UK.[95] This may mean that the service has a significant number of users in the UK, or the UK is a target market, or the service can simply be accessed in the UK.[96]

[84] Defamation Act 2013, s 5
[85] Giovanni De Gregorio, 'Democratising Online Content Moderation: a Constitutional Framework' (2020) Computer Law & Security Review 36(105374)
[86] Kate Klonick, 'The New Governors: The People, Rules and Processes Governing Online Speech' (2018) Harvard Law Review 131 1598
[87] Department of Digital, Culture, Media, and Sport, *Online Harms White Paper* (2020) <https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper> accessed 20 January 2023
[88] Joint Committee on the Draft Online Safety Bill, *Draft Online Safety Bill*, Report of Session 2021-2022.
[89] Department of Digital, Culture, Media, and Sport, *Online Harms White Paper* (2020) <https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper> accessed 20 January 2023
[90] ibid.
[91] Online Safety Act 2023, s 2.
[92] Online Safety Act 2023, s 2(1).
[93] Ashley Hurst and Ben Dunham, 'The UK Online Safety Bill Part 1: Is your Online Service and User Content within Scope?' (2021) Compliance & Risk 10(4)
[94] Ashley Hurst and Ben Dunham, 'The UK Online Safety Bill Part 1: Is your Online Service and User Content within Scope?' (2021) Compliance & Risk 10(4)
[95] Online Safety Act 2023, s 3(5) and s 3(6).
[96] Online Safety Act 2023, s 3(5) and s 3(6).

If the internet service is within the scope of the Act, the platform needs to consider what types of content are on its platforms and whether that means a duty of care is owed.[97] According to the Act, content that is regulated under it is any user-generated content in almost any form that is generated or uploaded by a user of that service and may then be encountered by others.[98] This is a wide definition, and because of this the Act attempts to clarify (and restrict) it partly by excluding certain types of content, such as emails and SMS messages.[99] Once an internet service has established that it and its content falls under the scope of the Act, the platforms are required to conform to certain duties of care.

The first duty of care concerns illegal content. Section 9(3) states that this is a duty to operate the service in such a way that it minimises the presence of illegal content, minimises the length of time this illegal content is present and that such content is swiftly taken down.[100] In addition to that, risk assessments must be taken that assess potential access to illegal content.[101] Furthermore, platforms have reporting and record-keeping obligations, as well as obligations to be transparent to their users as to what might be classified as illegal content.[102]

For content to be considered illegal under the Act, it must 'amount to a relevant offence.'[103] Relevant offences include, but are not limited to, straightforward offences such as terrorism or child sexual exploitation, but also less straightforward offences, such as harmful[104] and false[105] communications offences. These offences can be defined and widened by the Secretary of State.[106] (This is considered in further detail in the following section.)

Two other duties concern children-facing and adult-facing duties. All internet providers, as stated by section 10(3), have a duty to operate their system in such a way that it prevents children of any age from encountering, by means of service, primary priority content that is harmful to children.[107] In addition to that, Category 1 services[108] will have a duty to protect adults from priority content.[109] What exactly qualifies as primary priority content and priority content is as yet unknown, but will be defined in secondary legislation that is yet to be published.[110]

Internet services also have a duty to identify and remove harmful content of which they have 'reasonable grounds to believe that the nature of the content is such that there is a material risk of the content having, or indirectly having, a significant adverse or

[97] Ashley Hurst and Ben Dunham, 'The UK Online Safety Bill Part 1: Is your Online Service and User Content within Scope?' Compliance & Risk 10(4)

[98] Online Safety Act 2023, s 39.

[99] Online Safety Act 2023, s 39(2).

[100] Online Safety Act 2023, s 9(3).

[101] Online Safety Act 2023, s 9(6).

[102] Online Safety Act 2023, s 9(5).

[103] Online Safety Act 2023, s 41(2).

[104] Online Safety Act 2023, s 150.

[105] Online Safety Act 2023, s 151.

[106] Online Safety Act 2023, Part 6.

[107] Online Safety Act 2023, s 10(3).

[108] Online Safety Act 2023, s 11(2).

[109] And again, the Bill imposes risk assessments, transparency and reporting and redress system duties on such social media platforms).

[110] Ashley Hurst and Ben Dunham, 'The UK Online Safety Bill Part 2: Challenges for Service Providers in Implementing Obligations' (2021) Compliance & Risk 10(5)

psychological impact' on a child or adult of 'ordinary sensibilities.'[111] This, colloquially called 'legal but harmful content,'[112] is quite wide. After all, what is meant by 'having a significant adverse physical or psychological impact' is yet unknown, as is the definition of a child or adult of 'ordinary sensibilities.' Nevertheless, public statements indicate that these definitions are likely to include misinformation, disinformation, misogynistic abuse, harassment, material encouraging self-harm or eating disorders,[113] and even possibly covid-19 misinformation.[114]

In addition to the above, the Act also imposes a duty on Category 1 services to have regard to protecting users' freedom of expression and privacy,[115] as well as considering freedom of expression when making content-related decisions regarding journalistic content[116] and content of democratic importance.[117] Journalistic content appears to be intended as content that is generated for the purposes of journalism and is linked to the UK.[118] Content of democratic importance is defined as 'news publisher or regulated, user-generated content that must appear to be intended to contribute to democratic political debate in the UK.'[119] What exactly would appear to contribute to democratic political debate in the UK though, is unknown.[120]

How are these duties to be enforced? Ofcom[121] will have the power to fine companies up to £18 million, or 10 per cent of annual global turnover, whichever is higher, if they are failing in their duty of care.[122] Additionally, the Government's response to the White Paper also appears to empower Ofcom to impose criminal sanctions against individual executives or senior managers at technology firms if they, for example, do not respond in an accurate or timely manner to information requests by the regulator.[123]

That is not the only role Ofcom plays in the Online Safety Act. Ofcom will also issue a Code of Practice which will likely describe the ways in which content should be moderated, such as what algorithms can be used and how human moderators can moderate content,

---

111 Online Safety Act 2023, s 45-46.

112 Matthew MacLachlan, 'The Online Safety Bill – Broader Implications' (2022) Privacy & Data Protection 23(2)

113 Department for Digital, Culture, Media and Sport, *Online Safety Bill: Factsheet* (18 January 2023) <https://www.gov.uk/government/publications/online-safety-bill-supporting-documents/online-safety-bill-factsheet> accessed 20 January 2022

114 Matthew Lesh and Victoria Hewson, 'An Unsafe Bill: How the Online Safety Bill Threatens Free Speech, Innovation and Privacy' (2022) IEA Briefing Paper

115 Online Safety Act 2023, s 12(3).

116 Online Safety Act 2023, s 15.

117 Online Safety Act 2023, s 13.

118 Ashley Hurst and Ben Dunham, 'The UK Online Safety Bill Part 1: Is your Online Service and User Content within Scope?' Compliance & Risk 10(4)

119 Online Safety Act 2023, s 13.

120 Ashley Hurst and Ben Dunham, 'The UK Online Safety Bill Part 1: Is your Online Service and User Content within Scope?' Compliance & Risk 10(4)

121 Office of Communications in the UK. See generally https://www.ofcom.org.uk/online-safety/information-for-industry/roadmap-to-regulation Accessed 28 October 2023

122 Online Safety Act 2023, s 85(4).

123 Department of Digital, Culture, Media, and Sport, *Online Harms White Paper* (2020) <https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper> accessed 20 January 2023

as well as guidance on drawing a line between what content is harmful and what is not.[124] Nevertheless, what exactly will be in the Code of Practice is currently unknown.

4. TO WHAT EXTENT IS FREE SPEECH PROTECTED OR SACRIFICED – AN ANALYTICAL SUMMARY

The foregoing section on rationale and practices demonstrates that the UK Government intends to take significant steps to combat the online harms that are often highlighted as being of particular concern. And although the Act also imposes a duty on internet services to have regard to freedom of expression, one can ask whether free speech is really protected or instead, sacrificed, in the pursuit of the Government's agenda to combat online harms. This section explores these questions by (among other things) considering the theory behind free speech and how the four commonly cited arguments interact with relevant provisions of the Act. An opportunity is provided to critically examine parts of the Act that impact freedom of expression,[125] and relate to journalistic content[126] and to content of democratic importance.[127] By way of background discussion, there is some consideration of the four arguments so far as applying to situations of: 1) illegal content, and 2) harmful but legal content.

4.1 Illegal Content

As was seen in the above section, the Act aims to ensure that internet services remove illegal content.[128] Although most illegal content that amounts to a relevant offence currently stated in the Act is understandable, two issues arise for discussion.

First, some of the offences defined in the Act relate to 'harmful'[129] and 'false'[130] communications. The Act defines the 'harmful communications offence' as 'sending a message that is intended to cause at least serious distress to a likely audience,'[131] and the 'false communications offence' as 'sending information known to be false that causes non-trivial damage.'[132] One could ask, however, what classifies as 'serious distress' or 'non-trivial damage'. When considering the theory behind free speech, it is often argued that classifying something as harmful or false, might be a detriment to free speech. For example, the argument from truth, self-fulfilment and political participation posit that open discussion is important, and that open discussion cannot happen when certain opinions or statements are restricted.[133] Similarly, when arguing that one needs to be suspicious of the government, that includes being suspicious of the government classifying anything as harmful or even false.[134] Applying these to the provisions regarding harmful and false communications offences, one might see how that might become a problem, as the obscurity of the definitions might restrict speech that should not be restricted. In fact, as Trengove argues, one could also consider free speech legislation here, and ask how the obscurity of these

[124] Online Safety Act 2023, s 34.
[125] Online Safety Act 2023, s 12(3).
[126] Online Safety Act 2023, s 14.
[127] Online Safety Act 2023, s 13.
[128] Online Safety Act 2023, s 9.
[129] Online Safety Act 2023, s 150.
[130] Online Safety Act 2023, s 151.
[131] Online Safety Act 2023, s 150.
[132] Online Safety Act 2023, s 151.
[133] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 7-21.
[134] Frederick Schauer, *Free Speech: A Philosophical Enquiry* (Cambridge University Press 1982)

offences would justify the necessity and proportionality principle.[135] For example, internet services might remove content they classify as content that might cause 'serious distress' in fear of risking significant fines, and therefore over-remove content that is not necessary to be removed, raising the question whether the over-removal is really proportionate to free speech principles.[136]

Nevertheless, as discussed above, some restriction on speech is necessary to combat harm, even in the eyes of prominent philosophers,[137] prompting one to question what exactly qualifies as 'harmful' or 'false.' The idea of harm will be discussed in the following section.

A second issue that arises is that the Act would give the Secretary of State the power to widen the scope and determine what other offences might fall within this definition.[138] This power does not need to be further authorised or approved and therefore, as Trengove[139] points out, circumvents the important democratic mechanism of Parliamentary scrutiny. Without the right checks and balances in place, one can see how this could conflict with certain free speech principles as well, especially the two regarding political participation and suspicion of government.

For example, although the argument regarding political participation usually concerns the speaker, and whether the speaker can participate in political debate,[140] one can see how the powers the Secretary of State will be given could take that away: not only by not allowing for debate in Parliament in the first place, but also by making it possible for the Secretary of State to make certain discussions offences, including discussions that play a role in a political debate.

So too, does this provision interact with the argument that one needs to be suspicious of government, as one can see how giving too much power to one individual person might be controversial.[141] Nevertheless, this is often called the 'slippery slope' argument, suggesting that once too much power is given to one person, that person will gain more and more power and misuse his position. And, as is common knowledge, the 'slippery slope' argument often fails to come true, despite its previous warnings.

Therefore, one can question whether this provision offers sufficient protection to free speech, or whether free speech is instead sacrificed to pursue a government agenda. When considering that harmful and false communications may be criminalised, one could see how free speech might not necessarily be protected, as it is not entirely clear what this entail, and do not allow for open discussion. However, when it comes to the Secretary of State possibly abusing powers, the question is how likely that is to happen.

4.2 Harmful but Legal

What about 'harmful but legal'[142] content, which has since changed to 'illegal because it is harmful': do the provisions in the Act concerning this type of content still protect free speech sufficiently?

As stated in the previous section, under the Act, any content that social media platforms have 'reasonable grounds to believe that the nature of the content is such that there is a

[135] Markus Trengove et al, 'A Critical Review of the Online Safety Bill' (2022) Elsevier 3(8)
[136] ibid.
[137] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007)
[138] Online Safety Act 2023, Part 6.
[139] Markus Trengove et al, 'A Critical Review of the Online Safety Bill' (2022) Elsevier 3(8)
[140] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 18.
[141] Markus Trengove et al, 'A Critical Review of the Online Safety Bill' (2022) Elsevier 3(8)
[142] Online Safety Act 2023, s 45-46. Latest version replaces this with illegal because it is harmful.

material risk of the content having, or indirectly having, a significant adverse physical or psychological impact' on a child or adult of 'ordinary sensibilities,'[143] should be removed. As mentioned previously, this type of content has been classified by critics as 'harmful but legal'.[144]

It is here, that the author would like to draw attention to the question of what can be classified as harmful, a question that was also raised regarding the 'harmful' and 'false' communications offences, and what qualifies internet services and governments to constitute the definition of harm. Although the Government has stated that Ofcom will give guidance on this definition,[145] it is nevertheless not clear currently.

Mill's viewpoint might be a useful point of analysis here. Although Mill was generally in favour of open discussion to find the truth, Mill did admit that certain restrictions on speech were necessary to prevent harm.[146] In the words of Mill: 'the only purpose for which power can be rightfully exercised over any member of a civilised community … is to prevent harm to others'.[147] Nevertheless, he states that harm cannot arise from speech alone, but that something else must be present.[148] Here, he gives an example of a corn dealer, and argues that the opinion that corn dealers starve the poor should be unrestricted when circulated by the press, but may need restriction when delivered orally to an excited mob.[149] Thus, one could argue that statements on internet services should not be restricted, since they are generally not delivered to an excited mob. Nevertheless, this argument is not infallible: who is to say that a social media post for example, or a news article, does not cause an angry mob to form and cause harm to the corn dealer? Nevertheless, Mill strongly argues that 'there ought to exist the fullest liberty of pressing and discussion … any doctrine, however immoral it may be considered'.[150] Thus, he seems to have a narrow conception of harm,[151] meaning that, according to Mill, 'harmful but legal' content should probably be defined very narrowly.

Others, however, argue for an expansion of harm, such as Jeremy Waldron[152] and Ronald Dworkin,[153] who argue that the state should be allowed to regulate and punish hate speech, since hate speech undermines the human rights and dignity of others. However, scholars like Heinze[154] seem to disagree with that and argue that the state should allow at least some kind of harm to marginalised groups in the name of democracy. Thus, the question remains as to what exactly constitutes harm, and as to what harm should be allowed.

---

143 Online Safety Act 2023, s 45-46.

144 Matthew MacLachlan, 'The Online Safety Bill – Broader Implications' (2022) Privacy & Data Protection 23(2)

145 Department of Digital, Culture, Media, and Sport, *Online Harms White Paper* (2020) <https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper> accessed 20 January 2023

146 John Stuart Mill, *On Liberty* (2nd edn, J W Parker and Son 1859

147 John Stuart Mill, *On Liberty* (2nd edn, J W Parker and Son 1859) 88.

148 John Stuart Mill, *On Liberty* (2nd edn, J W Parker and Son 1859) 121.

149 ibid.

150 John Stuart Mill, *On Liberty* (2nd edn, J W Parker and Son 1859) 86.

151 Stefan Theil, 'Private Censorship and Structural Dominance: why Social Media Platforms should have Obligations to their Users under Freedom of Expression' (2022) Cambridge Law Journal 81(3)

152 Jeremy Waldron, *The Harm in Hate Speech* (Cambridge University Press 2012)

153 Ronald Dworkin, *Taking Rights Seriously* (Duckworth Books 1977)

154 Eric Heinze, *Hate Speech and Democratic Citizenship* (Oxford University Press 2016)

The same could be argued when one considers the argument that supports political participation. After all, the question again is what constitutes harm. Take for example, the 'Hunter Biden Laptop' story, where Twitter and Facebook suppressed a story regarding possible corruption by now US President, then Presidential Candidate, Joe Biden.[155] This was done based on possible misinformation,[156] even though it might have held certain substance, and could have cost Biden the election if circulated more. With this new Act, the Government could also classify such stories as harmful.

Similarly, one could ask whether content that argues against democracy classifies as harmful content or whether hate speech does.[157] After all, neither content is necessarily political speech, and could therefore be restricted if one would base their argument solely on free speech being important for political participation in a democratic society.

When considering the self-fulfilment argument, one can ask the same question. After all, this argument holds that all choices must be available to a person,[158] and taking away certain choices through content moderation because the content might be harmful, would not allow for all the choices being available. In fact, Scanlon argues that even harmful choices should be available to an individual.[159] Of course the question is whether free speech, and therefore a range of choices, leads to self-fulfilment, and if people are not able to live a happier life with the proposed content moderation by the Government. In Animal Defenders International[160] for example, it was held that a ban on paid political advertising on TV and radio was allowed to 'protect the democratic debate and process from distortion by powerful financial groups,'[161] showing the court did allow for some restrictions on access to information, even though that restriction was partly based on protecting the public from certain choices being more prevalent due to distortion by powerful financial groups.

Nevertheless, the argument that one needs to be suspicious of the government is particularly apposite here. As Theil argues, it seems like certain legislators even help internet services to have more power in deciding what is and is not allowed on their platforms, without having to take accountability.[162] In fact, in this case, it is not just the internet services that have a say on what should be allowed, but the Government as well. And as Schauer argues, any institution should not have that much power in deciding what is allowed to be said, regardless of any possible good intentions.[163]

---

[155] Freddy Gray, 'How Twitter Suppressed the Hunter Biden Laptop Story' (The Spectator, 3 December 2022) <https://www.spectator.co.uk/article/how-twitter-suppressed-the-hunter-biden-laptop-story/> accessed 29 January 2023; David Molloy, 'Zuckerberg Tells Rogan FBI Warning Prompted Biden Laptop Story Censorship' (BBC News, 26 August 2022) <https://www.bbc.co.uk/news/world-us-canada-62688532> accessed 29 January 2023

[156] ibid.

[157] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 19.

[158] Eric Barendt, *Freedom of Speech* (2nd edn, Oxford University Press 2007) 13.

[159] Thomas Scanlon, 'A Theory of Freedom of Expression' (1972) Philosophy and Public Affairs 204(1)

[160] *Animal Defenders International v United Kingdom* [2013] EMLR 28

[161] *Animal Defenders International v United Kingdom* [2013] EMLR 28, para 86.

[162] Stefan Theil, 'Private Censorship and Structural Dominance: why Social Media Platforms should have Obligations to their Users under Freedom of Expression' (2022) Cambridge Law Journal 81(3)

[163] Frederick Schauer, *Free Speech: A Philosophical Enquiry* (Cambridge University Press 1982)

Taking the above four arguments into account, it can be seen how free speech might be sacrificed instead of protected under this particular provision in the Act. Given that it is currently not clear what exactly classifies as 'harmful but legal,' there is a possibility that the provision is too wide, and instead allows platforms to remove content that should not be removed, simply because someone might find it harmful. Nevertheless, it will all depend on how the internet services and the Government treat this provision now that the Bill has become law.

4.3 Freedom of Expression, Journalistic Content and Democratic Importance

This section considers the duty on platforms to take freedom of expression,[164] journalistic content[165] and content of democratic importance[166] into account. Here, two important arguments are suggested: that of political participation and suspicion of government.

First, the Act only states that platforms should take the above 'into account.' One could therefore argue that not enough importance is paid to these issues.[167] After all, internet services could simply state that they thought about freedom of expression, for example, but decided to remove the content anyway, even though the content was not particularly harmful. If one considers one always needs to be suspicious of the government and powerful institutions, one could see how this potentially becomes a problem if social media platforms and the government seek to abuse their power.

Likewise, specifically when one considers internet services having to take content of democratic importance into account, one can ask what constitutes content of democratic importance.[168] As previously mentioned, would content that argues against democracy be of democratic importance? Similarly, would hate speech be of a democratic importance, specifically taking into account that it is not always certain what classifies as hate speech.[169] It would be helpful here, for example, to again draw attention to case law discussed above, such as the case of R v Scottow,[170] where it was held that discussions about transgender issues were types of discussions that lead to important political debate, and are therefore not necessarily speech that should be restricted, even though some might classify it as harmful.

Thus, one can ask whether the relevant provisions of the Online Safety Act genuinely do enough to protect free speech. Although simply having to 'take into account' these provisions it might appear that the Act assigns more importance to the combatting of online harms. It all depends on how the Act will be implemented, both by the internet services and the Government, something that is currently hard to tell given the recency of the Bill becoming law.

5. CONCLUSION

Here again is the central question: does the Online Safety Act protect free speech? If so, to what extent does it do that? Little is yet known on what the proposed Code of Practice

[164] Online Safety Act 2023, s 12(3).

[165] Online Safety Act 2023, s 14.

[166] Online Safety Act 2023, s 13.

[167] Peter Coe, 'Misinformation, Disinformation, the Online Safety Bill, and its Insidious Implications for Free Speech' (2021) Communications Law 26(3)

[168] Alexander Dittel, 'The UK's Online Safety Bill: the Day we took a Stand against Serious Online Harms or the Day we Lost our Freedoms to Platforms and the State?' (2022) Journal of Data Protection and Privacy 5(2)

[169] Peter Coe, 'The Draft Online Safety Bill and the Regulation of Hate Speech: Have we Opened Pandora's Box?' (2022) Journal of Media Law 14(1)

[170] *R v Scottow* [2020] EWHC 3421

will look like, as (at the time of writing) it has not been published. Although it is known that the Ofcom Code of Practice will likely describe ways in which content should be moderated, little is known about the ways it will approve of. For example, would it advocate for AI algorithms and machine learning techniques, or would it place a heavier focus on human content moderators? Similarly, although it is known that it will likely offer guidance on what should be classified as 'harmful content,' little is currently known as to how exactly the Code of Practice will define it. Because of this, no careful analysis is possible as to whether the Code of Practice will do enough to protect free speech.

In its design of the passed Bill, the Government's main aim is to combat online harms, such as child sexual exploitation and terrorist content; but the aim is also to address more obscure harms such as the 'harmful communications offence,' the 'false communications offence' and the removal of 'legal but harmful' content. In examining relevant provisions of the Act, this article considered the four commonly cited free speech arguments and explored how each might interact with the provisions of the Act. It was seen, for example, that provisions authorising or requiring the removal of illegal content, might run counter to certain free speech arguments – such as the arguments from political participation and suspicion of government, due to the obscurity of certain definitions and the powers it gives to the Secretary of State. Likewise, it was seen that the obscurity of the definition of 'legal but harmful' content, might conflict with every free speech argument, although it depends on how 'harm' will be classified and to what extent it will be allowed.

It is concluded, more generally, that, given that certain provisions of the Online Safety Act 2023 merely put a duty on social media platforms to 'take into account' free speech, and place more importance on provisions that impose duties concerning illegal and 'harmful but legal' content, freedom of expression is not explicitly protected. In fact, the main concerns of the UK Government, namely the online harms, seem to be squarely addressed by the provisions of the Act, whereas free speech comes across as simply an afterthought. In other words, although the intent of the Government to fight online harms is accomplished by the provisions, free speech might be sacrificed in furtherance of the Government's agenda.

Protection of free speech would depend on how Ofcom, the regulator and competition authority for the communication industries, and internet services both interpret and implement the Online Safety Act, when the former issues its code of practice. Future studies on the Act and the attendant Ofcom Code of Practice could therefore consider a comparative engagement with activities in the European Union. For example, the EU Digital Services Act[171] that replaces its E-Commerce Directive could be compared with the Online Safety Act.

Acknowledgement:

This interdisciplinary study in Human Rights, Legal Theory and the Media has benefited immensely from insights shared by a host of critical friends and members of the International Forum for the Study of Jurisprudence and Value Inquiry. The researchers are very grateful to the anonymous peer reviewers for their detailed and helpful comments. Shekhinah and Seraphina, for being there both to inspire and to encourage during the final phase of the write up, here is a short note to say a big THANK YOU!

[171] See generally The Digital Services Act: ensuring a safe and accountable online environment Available at The EU's Digital Services Act (europa.eu) Accessed 31 October 2023

Bibliography/References

1. Access Now et al., Santa Clara Principles on Transparency and Accountability in Content Moderation (The Santa Clara Principles, 2018) <https://santaclaraprinciples.org/> accessed 24 March 2023
2. Alexander L and Horton P, 'The Impossibility of a Free Speech Principle' (1983) North Western University Law Review 78(1319)
3. Baker E, Human Liberty and Freedom of Speech (Oxford University Press 1989)
4. Belli L, Francisco P and Zingales N, 'Law of the Land or Law of the Platform? Beware of the Privatisation of Regulation and Police' in Belli L and Zingales N (eds), How Platforms are Regulated and how They Regulate Us (FGV Direito Rio, 2017)
5. Balkin J M, 'Digital Speech and Democratic Culture: a Theory of Freedom of Expression for the Information Society' (2004) New York University Law Review 79(1)
6. Barendt E, Freedom of Speech (2nd edn, Oxford University Press 2007)
7. Berger J M and Perez H, 'The Islamic State's Diminishing Returns on Twitter: How Suspensions are Limiting the Social Networks of English-Speaking ISIS Supporters' (2016) Program of Extremism
8. Bork R, 'Neutral Principles and Some First Amendment Problems' (1971) Indiana Law Journal 47(1)
9. Boyd D and Ellison N, 'Social Network Sites: Definition, History, and Scholarship' (2007) Journal of Computer-Mediated Communication 13(1)
10. Bruns A, 'Making Sense of Society Through Social Media' (2015) Social Media + Society 1(1)
11. Capurso T J, 'How Judges Judge: Theories on Judicial Decision Making' (1998) University of Baltimore Law Forum 29(1)
12. Carr C and Hayes R, 'Social Media: Defining, Developing, and Divining' (2015) Atlantic Journal of Communication 23(1)
13. Clark L, 'AI is being Used to Hunt out Child Porn and Sexual Abuse Images across the Web' (12dJanuaryd2016,dWired) <https://www.wired.co.uk/article/ai-interpol-track-child-abuse> accessed 20 January 2023
14. Clark M and Weatherbed J, 'YouTube Creators are Ducking Outraged by Swearing Policy' (13 January 2023, The Verge) <https://www.theverge.com/2023/1/13/23553746/youtube-swearing-advertising-policy-change> accessed 20 January 2023
15. Coe P, 'Misinformation, Disinformation, the Online Safety Bill and its Insidious Implications for Free Speech' (2021) Communications Law 26(3)
16. Conger K and Hirsch L, 'Elon Musk Completes $44 Billion Deal to Own Twitter' (New York Times, 27 October 2022) <https://www.nytimes.com/2022/10/27/technology/elon-musk-twitter-deal-complete.html> accessed 23 March 2023
17. Crawford K and Gillespie T, 'What is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint' (2016) New Media & Society 410
18. Dean B, 'Social Network Usage & Growth Statistics: how many People use Social Media in 2022?' (Backlinko, 10 October 2021) <https://backlinko.com/social-media-users> accessed 29 January 2023
19. De Gregorio G, 'Democratising Online Content Moderation: a Constitutional Framework' (2020) Computer Law & Security Review 36(105374)
20. Demangeot C and Broderick A J, 'Consumer Perceptions of Online Shopping Environments: A Gestalt Approach' (2010) Psychology and Marketing 27(2)

21. Department of Digital, Culture, Media, and Sport, Online Harms White Paper (2020) <https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper> accessed 20 January 2023
22. Department for Digital, Culture, Media and Sport, Online Safety Bill: Factsheet (18dJanuaryd2023) <https://www.gov.uk/government/publications/online-safety-bill-supporting-documents/online-safety-bill-factsheet> accessed 20 January 2022
23. Dittel A, 'The UK's Online Safety Bill: the Day we took a Stand against Serious Online Harms or the Day we Lost our Freedoms to Platforms and the State?' (2022) Journal of Data Protection and Privacy 5(2)
24. Dworkin R, 'Is there a Right to Pornography' (1981) Oxford Journal of Legal Studies 177(1)
25. Dworkin R, Taking Rights Seriously (Duckworth Books 1977)
26. Eecke P and Capdevila E, 'The Proposed Digital Services Act Packages – what you Need to Know' (2021) Privacy and Data Protection 21(3)
27. English Language Learning, Cambridge Dictionary (4th edn, Cambridge University Press, 2012)
28. Fisher M, 'Inside Facebook's Secret Rulebook for Global Political Speech' (27dDecemberd2018,dThedNewdYorkdTimes) <https://www.nytimes.com/2018/12/27/world/facebook-moderators.html> accessed 20 January 2023
29. Flew T, Martin F and Suzor N, 'Internet Regulation as Media Policy: Rethinking the Question of Digital Communication Platform Governance' (2019) Journal of Digital Media & Policy 10(1)
30. Fuchs C, Social Media: a Critical Introduction (2nd edn, Sage Publishers 2017)
31. Gadde V, 'Twitter Executive: Here's How We're Trying to Stop Abuse while Preserving Free Speech' (The Washington Post, 16 April 2015) <https://www.washingtonpost.com/posteverything/wp/2015/04/16/twitter-executive-heres-how-were-trying-to-stop-abuse-while-preserving-free-speech/> accessed 13 July 2022
32. Ghani N A, Hamid S, Hashem I A T and Ejaz Ahmed, 'Social Media Big Data Analytics: A Survey' (2019) Computers in Human Behaviour 101
33. Gorwa R, Binns R and Katzenbach C, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) Big Data & Society 7(1)
34. Gray F, 'How Twitter Suppressed the Hunter Biden Laptop Story' (The Spectator, 3 December 2022) <https://www.spectator.co.uk/article/how-twitter-suppressed-the-hunter-biden-laptop-story/> accessed 29 January 2023
35. Hall J, 'When is Social Media Use Social Interaction? Defining Mediated Social Interaction' (2016) New Media and Society 20(1)
36. Heinze E, Hate Speech and Democratic Citizenship (Oxford University Press 2016)
37. Hern A, 'Journalists, Politicians and Judges to Sit on Facebook's Free Speech Panel' (6 May 2020, The Guardian) <https://www.theguardian.com/technology/2020/may/06/facebook-oversight-board-freedom-expression-helle-thorning-schmidt-alan-rusbridger> accessed 24 March 2023
38. Hogan B and Quan-Haase A, 'Persistence and Change in Social Media' (2010) Bulletin of Science, Technology & Society 30(5)
39. Howard P, Duffy A, Freelon D, Hussain M, Mari W and Maziad M, 'Opening Closed Regimes: What was the Role of Social Media During the Arab Spring?' (2015) Social Media and Authoritarianism 1(1)

40. Hurst A and Dunham B, 'The UK Online Safety Bill Part 1: Is your Online Service and User Content within Scope?' Compliance & Risk 10(4)
41. Hurst A and Dunham B, 'The UK Online Safety Bill Part 2: Challenges for Service Providers in Implementing Obligations' (2021) Compliance & Risk 10(5)
42. Ingram M, 'Here's Why Facebook Removing that Vietnam War Photo is so Important'd(9dSeptemberd2016,dFortune) <https://fortune.com/2016/09/09/facebook-napalm-photo-vietnam-war/> accessed 20 January 2023
43. Joint Committee on the Draft Online Safety Bill, Draft Online Safety Bill, Report of Session 2021-2022.
44. Klonick K, 'The New Governors: The People, Rules and Processes Governing Online Speech' (2018) Harvard Law Review 131(1598)
45. Koltay A, 'Constitutional Protection of Lies?' (2020) Communications Law 25(3)
46. Kraus R, 'The Public Doesn't Agree with Elon Musk's Freedom of Speech Twitter Crusade' (Mashable, 15 April 2022) <https://mashable.com/article/elon-musk-freedom-of-speech-public-opinion-disconnect> accessed 23 March 2023
47. Krishnan N, Gu J, Tromble R and Abroms L, 'Research Note: Examining how Various Social Media Platforms have Responded to Covid-19 Misinformation' (2021) Harvard Kennedy School Misinformation Review 2(6)
48. Langvardt K, 'Regulating Online Content Moderation' (2018) The Georgetown Law Review 106(1353)
49. Lau W, 'Effects of Social Media Usage and Social Media Multitasking on the Academic Performance of University Students' (2017) Computers in Human Behaviour 68
50. Lazarus R and Folkman S, Stress, Appraisal and Coping (New York Springer 1999)
51. Lesh M and Hewson V, 'An Unsafe Bill: How the Online Safety Bill Threatens Free Speech, Innovation and Privacy' (2022) IEA Briefing Paper
52. Loftus E and Palmer J, 'Reconstruction of Auto-Mobile Destruction: an Example of the Interaction between Language and Memory' (1974) Journal of Verbal Learning and Verbal Behaviour 13
53. MacLachlan M, 'The Online Safety Bill – Broader Implications' (2022) Privacy & Data Protection 23(2)
54. Malice M, The White Pill: A Tale of Good and Evil (Independently Published, 2022)
55. McNamara H, 'Four Problems with Twitter's Irresponsible Policy on Pornography' (National Center on Sexual Exploitation, 8 January 2020) <https://endsexualexploitation.org/articles/four-problems-with-twitters-irresponsible-policy-on-pornography/> accessed 29 January 2023
56. Mangold G and Faulds D, 'Social Media: The New Hybrid Element of the Promotion Mix' (2009) Journal of Business Horizons 52
57. McCay-Peet L and Quan-Haase A, 'What is Social Media and What Can Social Media Research Help us Answer?' in The SAGE Handbook of Social Media Research Methods (SAGE Publications Ltd 2016)
58. McCloskey H J, 'Liberty of Expression: its Grounds and Limits' (2008) An Interdisciplinary Journal of Philosophy 13(1)
59. Meiklejohn A, Free speech and its Relation to Self-Government (Harper 1948)
60. Meiklejohn A, 'The First Amendment is an Absolute' (1961) Supreme Court Review 245
61. Merriam Webster, Definition of Threat (Merriam Webster) <https://www.merriam-webster.com/dictionary/threat> accessed 29 January 2023
62. Meta Transparency Center, 'Facebook Community Standards' (Meta)

<https://transparency.fb.com/en-gb/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards> accessed 24 March 2023
63. Mill J S, On Liberty (2nd edn, J W Parker and Son 1859)
64. Molloy D, 'Zuckerberg Tells Rogan FBI Warning Prompted Biden Laptop Story Censorship' (BBC News, 26 August 2022) <https://www.bbc.co.uk/news/world-us-canada-62688532> accessed 29 January 2023
65. Monat A and Lazarus R, 'Stress and Coping: Some Current Issues and Controversies' in Alan Monat and Richard Lazarus (eds) Stress and Coping: an Anthology (Columbia University Press 1991)
66. Moore V, 'Free Speech and the Right to Self- Realisation' (2005) UCL Jurisprudence Review 12
67. Morrison E, 'The Lab Leak and other Conspiracy Theories, Some of Which Turned out to be True' (Areo, 28 May 2021) <https://areomagazine.com/2021/05/28/the-lab-leak-and-other-conspiracy-theories-many-of-which-turned-out-to-be-true/> accessed 29 January 2023
Musk E, 'Tweet: Free Speech is Essential to a Functioning Democracy. Do you Believe Twitter Rigorously Adheres to this Principle?' (Twitter, 25 March 2022) <https://twitter.com/elonmusk/status/1507259709224632344?lang=en> accessed 23 March 2023
68. Oozeer A, 'Internet and Social Networks: Freedom of Expression in the Digital Age' (2014) Commonwealth Law Bulletin 40(2)
69. Oxford Languages, Oxford Dictionary of English (3rd edn, Oxford University Press 2010)
70. Pagana K, 'Stressed and Threats Reported by Baccalaureate Students in Relation to an Initial Clinical Experience' (1988) Journal of Education 27
71. Papacharissi Z, 'We have Always been Social' (2015) Social Media + Society 1(1)
72. Pasquale F, 'Platform Neutrality: Enhancing Freedom of Expression in Spheres of Private Power' (2016) Theoretical Inquiries L. 17
73. Rhee L, Bayer J B, Lee D S and Kuru O, 'Social by Definition: How Users Define Social Platforms and Why it Matters' (2021) Telematics and Informatics 59
74. Roberts S, 'Digital Detritus: "Error" and the Logic of Opacity in Social Media Content Moderation' (2018) First Monday 23(3)
75. Rosen J, 'Elon Musk is Right that Twitter Should Follow the First Amendment' (The Atlantic, 2 May 2022) <https://www.theatlantic.com/ideas/archive/2022/05/elon-musk-twitter-free-speech-first-amendment/629721/> accessed 23 March 2023
76. Rowbottom J, 'To Rant, Vent and Converse: Protecting Low Level Digital Speech' (2012) Cambridge Law Journal 71(2)
77. Sander B, 'Democratic Disruption in the Age of Social Media: between Marketized and Structural Conceptions of Human Rights Law' (2021) European Journal of International Law 32(1)
78. Scanlon T, 'A Theory of Freedom of Expression' (1972) Philosophy and Public Affairs 204(1)
79. Schauer F, Free Speech: A Philosophical Enquiry (Cambridge University Press 1982)
80. Scholtz S, 'Threat: Concept Analysis' (2000) N Forum 35(4)
81. Schwiddessen S, Clark B, Defaux T and Groom J, 'Germany's Network Enforcement Act – Closing the Net on Fake News?' (2018) European Intellectual Property Review 40(8)

82. Siddiqui F and Dwoskin E, 'Elon Musk Acquires Twitter and Fires Top Executives' (The Washington Post, 28 October 2022) <https://www.washingtonpost.com/technology/2022/10/27/twitter-elon-musk/> accessed 23 March 2023
83. Smith K, '126 Amazing Social Media Statistics and Facts' (Brandwatch, 30 December 2019) <https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/> accessed 8 November 2022
84. Stempel J and Frankel A, 'Twitter Sued by United States Widow for Giving Voice to Islamic State' (Reuters, 15 January 2016) <https://www.reuters.com/article/us-twitter-isis-lawsuit-idUSKCN0US1TA> accessed 29 January 2023
85. Theil S, 'Private Censorship and Structural Dominance: why Social Media Platforms should have Obligations to their Users under Freedom of Expression' (2022) Cambridge Law Journal 81(3)
86. Towey H, 'Elon Musk Now Owns Twitter. Here are the Busy Billionaire's 4 other Companies and what They all Do.' (Business Insider,
87. 28 October 2022) <https://www.businessinsider.com/elon-musk-companies-tesla-spacex-boring-co-neuralink-twitter-2022-4?r=US&IR=T> accessed 23 March 2023
88. Trengove M et al, 'A Critical Review of the Online Safety Bill' (2022) Elsevier 3(8)
89. Tufekci Z, 'Youtube, the Great Radicalizer' (The New York Times, 10 March 2018) <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html> accessed 29 January 2023
90. Turner C, 'Teacher Accused of 'Misgendering' Child was Told by the Police that she Committed a Hate Crime' (The Telegraph, 23 February 2018) <https://www.telegraph.co.uk/education/2018/02/23/teacher-accused-misgendering-child-told-police-committed-hate/> accessed 29 January 2023
91. Twitter Help Center, 'The Twitter Rules' (Twitter) <https://help.twitter.com/en/rules-and-policies/twitter-rules> accessed 24 March 2023
92. Waldron J, The Harm in Hate Speech (Cambridge University Press 2012)
93. Wertz J, 'How Social Values Drive Consumers to Brands' (27 December 2021, Forbes) <https://www.forbes.com/sites/jiawertz/2021/12/27/how-social-values-drive-consumers-to-brands/?sh=170990c27425> accessed 20 January 2023
94. Williams B, Truth and Truthfulness: an Essay in Genealogy (Princeton University Press 2004)
95. Wragg P, 'Mill's Dead Dogma: The Value of Truth to Free Speech Jurisprudence' (2013) Public Law