Northern Ireland
Legal Quarterly
Vol. 76 No. AD1 (2025) 83–135
Articles DOI: 10.53386/nilq.v76iAD1.1217

The Online Safety Act 2023: fostering democratic participation while combatting anti-democratic harms?

Helen Fenwick
Durham University

Peter Coe
University of Birmingham*
Correspondence emails: h.m.fenwick@durham.ac.uk; p.j.coe@bham.ac.uk.

ABSTRACT

This article is the first to single out and analyse one specific aspect of the Online Safety Act 2023 (OSA): its ability to combat anti-democratic online content and practices. The original perception of the internet as an egalitarian, democratic, expression-rich environment, free of burdensome regulation and of the dominance of global 'traditional' media companies, has given way to a focus on the harms its 'lawless' nature are deemed to create. The idea that the availability of online platforms fosters free expression and therefore promotes the health of democracies is coming into conflict with concerns as to the anti-democratic impact of some online practices and content. The spread of false information online, as destabilising the democratic process and undermining faith in elections, is far from the only concern, but it is a highly dominant one. As a result, intense pressure has been placed on governments globally to combat such anti-democratic tendencies, largely via regulation of online content. That pressure was one of the driving factors behind the introduction of the OSA in the United Kingdom. It was presented as creating a new model of sanitising internet governance, able, *inter alia*, to address online harms inimical to the health of democracy while preserving content of democratic importance. This is the first article to question its success in that venture, by interrogating the OSA, mainly in terms of its ability to create curbs on anti-democratic online content and practice (in particular, algorithmic tendencies), while also demonstrating efficacy in navigating the tensions between the creation of such curbs and the promotion of such content.

Keywords: Online Safety Act 2023; online harms; criminal law; free expression; democracy; regulation.

* Helen Fenwick is Professor of Law, School of Law, Durham University; Peter Coe is Associate Professor in Law, Birmingham Law School, University of Birmingham. We are indebted to Northern Ireland Legal Quarterly's anonymous reviewers for their invaluable comments on an earlier draft of this article.

INTRODUCTION

The specific focus of this article is on online harms generally viewed as tending to undermine the healthy workings of democracy and the failures of self-regulation by the tech companies in curbing them. The Online Safety Act 2023 (OSA) represents an attempt at introducing non-voluntary regulation, aimed at addressing a range of online harms more effectively by imposing statutory duties on companies to curb them, which will be monitored by Ofcom. This article is the first to single out and analyse one specific but highly significant aspect of the OSA: its ability to combat anti-democratic online harms. In probing the new scheme's success in that venture, this article develops its thesis as to both the anti-democratic and pro-democratic impacts of online content, critiquing the ways in which the OSA scheme attempts to enable a more effective navigation between the two than self-regulation achieved. While this article thus focuses specifically on the OSA's role in relation to combatting anti-democratic online harms, several of the points made below as to the general nature of the scheme created would also apply to its combatting, or failures to combat, other online harms.

The dereliction of the companies' responsibilities to their users in terms of failing to combat anti-democratic online harms is traced in the first section of this article, below. It has been compounded by the inadequacies of the laws, quasi-legal and non-legal measures, that were in place for dealing with illegal and harmful content, but largely designed for the offline analogue world, and were therefore outdated and no longer fit for purpose.[1] A number of such laws, in any event, were aimed at the creator of the content, not the hosting service. The key, general means, aside from data protection curbs, of monitoring and controlling online content – self-regulation – is viewed as having failed to prevent online harms, including anti-democratic ones. As a result, self-regulation is giving way in a number of instances to top-

1 Dame Melanie Dawes (Ofcom), 'Keynote speech: In news we trust: keeping faith in the future of media' (Oxford Media Convention 19 July 2021).

down regulation by a regulator, globally.[2] In line with this global effort, the new United Kingdom (UK) online safety regime under the OSA was eventually introduced, after an exceptionally lengthy, convoluted process,[3] providing many opportunities for divergent views of ministers to emerge and for tech company lobbying of Members of Parliament (MPs) and officials.[4] It purports to represent a solution to this problem by presenting a new model of governance of online platforms and services, intended to combat a range of online harms, including apparently the anti-democratic tendencies of some online content. Possibly the initial determination to link the OSA to the promotion of democracy was watered down by the Conservative Government during the lengthy process from *White Paper* to royal assent, but it is still apparent as an aim, not least in the provisions demanding the promotion online of 'content of democratic importance'.[5] The OSA does *not* represent an attempt to end self-regulation entirely, and such regulation will continue, but the intention is that it will be subject to

2 The key example is the EU's Digital Services Act 2022 (DSA). See also Ireland's Online Safety and Media Regulation Act 2022, Germany's Network Enforcement Act 2017 (Netzwerkdurchsetzungsgesetz) (NetzDG), Singapore's Protection from Online Falsehoods and Manipulation Act 2019, Australia's Online Safety Act 2021, Sri Lanka's Online Safety Act 2024. For discussion, see M Husovec, 'Rising above liability: the Digital Services Act as a blueprint for the second generation of global internet rules' (2023) 38(3) Berkeley Technology Law Journal 882–920; S Maaß et al, 'Evaluating the regulation of social media: An empirical study of the German NetzDG and Facebook' (2024) 48(5) Telecommunications Policy 102719.

3 In 2019, general, non-voluntary regulation of online expression entered the parliamentary agenda in the form of the *Online Harms White Paper* (hereafter referred to as *White Paper*): HM Government, *Online Harms White Paper*, CP 57, April 2019. See also *Online Harms White Paper: Full Government Response to the Consultation*, CP 354, December 2020. The *White Paper* was preceded, in October 2017, by a Department for Digital, Culture, Media and Sport green paper titled 'Internet Safety Strategy'. This transmuted into multiple iterations of the Online Safety Bill (OSB), published in its original form in May 2021. The OSA received royal assent on 26 October 2023.

4 Four Prime Ministers and five Digital Ministers have taken varying stances on the development of the legislation since the proposals were first published. See M Scott and A Dickson, 'How UK's Online Safety Bill fell victim to never-ending political crisis' (*Politico* 28 February 2023).

5 S17 OSA. A key aim put forward originally by the UK Government was to address anti-democratic online harms; it was reiterated throughout the *White Paper* (see n 3 above). For example, see Executive Summary, para 2; para 4; pt 1, paras 1.22–1.24 and particularly boxes 13 and 14; pt 3, para 7.25. See also Department for Digital, Culture, Media & Sport and Home Office, 'Landmark laws to keep children safe, stop racial hate and protect democracy online published' (Gov.uk 12 May 2021). The Government's press release on the Bill, when introduced into Parliament stated: 'It will also put requirements on social media firms to protect journalism and democratic political debate on their platforms' (17 March 2022).

interventions under the OSA scheme.[6] Indeed, the OSA may *encourage* the continuance of self-regulation[7] since the regulated services are likely to continue to rely on their own codes, possibly modified, in order to avoid Ofcom's interventions.

But, *prima facie*, the new OSA regime ends 'the era of self-regulation',[8] as the key form of *general* regulation of the tech companies, by creating a partially top-down regulatory system that imposes responsibility on the platforms themselves through the imposition of statutory safety duties of care to protect users from certain illegal content and, in the case of under-18s, from some harmful but legal content.[9] At its core, the regime is risk-based: regulated services are required to conduct risk assessments of services regarding illegal content,[10] as well as content harmful to children, and to implement effective and proportionate risk-mitigation plans in relation to the design or operation of the service in response.[11] The inception of the OSA means, if the regulatory scheme is taken at face value, that for the first-time in-scope services, as intermediaries, can be subjected to sanctions based on enabling illegal content to be made accessible via the service, published or made available by third parties. Despite the basis for its introduction, the OSA, the *White Paper*, and the Bill attracted opprobrium from a range of critics, including from pro- and anti-democratic opposing viewpoints.[12] It has been argued that the scheme fails to hold the services it regulates sufficiently to account, still leaving them leeway to perpetrate and enable the proliferation of anti-democratic cyber-harms, particularly by way of promulgation

6 The *White Paper* found that the existing 'patchwork of regulation and voluntary initiatives' was not effective in keeping users safe online, necessitating the imposition of a single regulatory framework: (n 3 above) 6, para 7, and 30.

7 Ibid. The *White Paper* proposed that self-regulation would continue alongside the new regime: para 2.10, 35.

8 Lord Bishop of Oxford, HL Deb 18 May 2021, vol 812, col 517.

9 In the OSA, the safety duties relating to adults are set out at s 10 (user-to-user services) and s 27 (search services) and for children at ss 12 and 29 (user-to-user and search services respectively). Obviously, online content is subject to laws aimed at curbing expression, including defamation law, but often, as discussed below, aside from data protection provisions (in particular, the UK General Data Protection Regulation), they tend to be aimed at the person creating the content, not at the platform hosting it.

10 OSA ss 9, 26.

11 See ss 9, 10, 23 (user-to-user services) and ss 26 and 27 (search services). These provisions determine assessment duties and resultant safety duties.

12 In Autumn 2022, it appeared that the Bill, and the entire regime, was on the verge of being abandoned altogether. See also H Schmidt, 'The Online Safety Act 2023' (2024) 16(2) Journal of Media Law 202–210, 205.

of user-generated false information.[13] But, conversely, other actors have criticised it, and top-down regulation generally, as failing in pro-democratic terms due to incursions into free speech and media freedom, based on the encouragement of online censorship by the services on market-based grounds.[14] This article, due to its particular focus, critiques the OSA from both viewpoints, commenting on its attempts to preserve free expression online, thereby furthering democratic ends,[15] while minimising online harms inimical to the healthy functioning of democracies. As indicated above, its key concern is with the ability of this model of regulation to navigate a path between fostering the pro-democratic potential of online content while curbing such harms.

It will be claimed that the OSA is readily open to criticism from two opposing viewpoints, namely that it fails to further democratic ends in several respects, while largely failing to curb anti-democratic harms. We acknowledge at the outset that achieving those two conflicting aims simultaneously is a somewhat daunting task for any model of online regulatory legislation, but the key argument of this article is that the OSA is not suitable for this arduous task and falls short of meeting those aims. The problem lies, not with the aims – which in our view, such

13 Eg E Abrusci, 'The UK Online Safety Act, the EU Digital Services Act and online disinformation: is the right to political participation adequately protected?' (2024) 16(2) Journal of Media Law 1–28, 17–28; A Hern, 'Why Musk's rabble-rousing shows the limits of social media laws' *The Guardian* (London 13 August 2024); V Pickard, writing from a US perspective, examines the general pro-democratic basis for regulating online content, 'A new social contract for platforms' (ch 17), and D Tambini, 'Reconceptualising media freedom' (ch 16) both in D Tambini and M Moore (eds), *Regulating Big Tech* (Oxford University Press 2022). These criticisms are often predicated on the argument that the previous model based on self-regulation alone presented a misalignment between market incentives and the fostering of democratic ends.

14 Eg Index on Censorship, Right to Type (4 June 2021); House of Lords Communications and Digital Committee, *Free for All? Freedom of Expression in the Digital Age*, First Report of Session 2021–22, HL Paper 54, 22 July 2021; M Earp, 'UK Online Safety Bill raises censorship concerns and questions on future of encryption' (Committee to Protect Journalists 25 May 2021); S Dawood, 'Will the Online Safety Act protect us or infringe our freedoms?' (*The New Statesman* 17 November 2023).

15 Eg A Bhagwat and J Weinstein, 'Freedom of expression and democracy' in A Stone and F Schauer (eds), *The Oxford Handbook of Freedom of Speech* (Oxford University Press 2021) ch 5; R Post, 'Democracy and equality' (2006) 603(1) Annals of the American Academy of Political and Social Science 24–36; K Greenawalt, 'Free speech justifications' (1989) 89 Columbia Law Review 119–155, 143; V Blasi, 'The checking value in First Amendment theory' (1977) 2(3) American Bar Foundation Research Journal 521–649. While these works concern free speech offline, their messages are also applicable to online expression.

regulatory schemes are under an expectation of seeking to satisfy[16] – but largely with its execution via the drafting of the OSA. Admittedly, the difficulties are, to an extent, inevitable where tech companies are required to moderate vast quantities of content by making qualitative judgements about it, but the specific criticisms of the OSA advanced below are intended to demonstrate that this particular scheme has greatly exacerbated them.

To that end, it begins in the next, first, section by exploring the challenges faced by self-regulation, and now by the OSA regulatory scheme, in seeking to preserve the speech-based pro-democratic benefits of online services, while addressing the anti-democratic harms they also create. The nature of a range of such harms, using three key examples, is also analysed in that section, in order to revisit them in the following, second, section, contrasting the new OSA regulation with reliance on self-regulation in navigating a path between the pro- and anti-democratic concerns viewed here as at stake. The second section, as it details in its introduction, considers the methods utilised by the OSA to tackle such harms and identifies three specific weaknesses in the OSA scheme which mean that is unlikely to be effective in tackling them. It then considers those weaknesses in relation to the three key examples of those harms considered in the first section. Then, in its third and final section the article explores further inherent weaknesses in the OSA regulatory scheme, which, it will be argued, may indicate that aspirations to enable it to navigate an effective path between curbing the anti-democratic impacts of online services and preserving free expression were never wholeheartedly embraced: the market freedom of the companies still appears to be the priority. The weaknesses identified are not confined to failures in relation to such impacts; they also apply to combatting other online harms: they range from raising doubts as to the ability of the regulator, Ofcom, to enforce the scheme, to an analysis of the influence on it of governmental figures who are susceptible to lobbying and pressure from the tech companies.

16 This is supported by Tambini and Moore who argue that: 'It has slowly dawned on citizens of democracies that … democratic decision-making … will be increasingly compromised if the digital status quo [self-regulation by the tech companies] is allowed to continue.' See Tambini and Moore (n 13 above) 1.

PRO- AND ANTI-DEMOCRATIC PROPENSITIES OF ONLINE CONTENT: FAILURES OF SELF-REGULATION

Expanding and enriching the public sphere: democratising expression

The pro-democratic capacities of the internet, seen by some as the 'Fifth Estate',[17] are indisputable: it provides the technological architecture for greater egalitarian engagement with the public sphere, by allowing individuals to circumvent institutional, financial and technological barriers to communication,[18] news production and consumption.[19] Its *potential* therefore for democratising and enriching the public sphere by accommodating a greater diversity of voices, ideas and information is readily apparent.[20] This section begins, therefore, by evaluating the sense in which online services make a particular contribution to supporting democratic processes. The impulse of individuals to communicate and engage in debate with others, highly relevant to serving democratic aims, is clearly more fully facilitated than via offline methods. That determination to communicate is matched and underpinned by the ready availability of online services.[21] In becoming an indispensable part of everyday life,[22] online content has, as van Dijck finds, 'penetrated every fibre of culture today' by

17 W H Dutton, 'The Fifth Estate emerging through the network of networks' (2009) 27(1) Prometheus 1–15.

18 A Koltay, *New Media and Freedom of Expression: Rethinking the Constitutional Foundations of the Public Sphere* (Hart 2019) 74; B Wellman, 'Physical space and cyberspace: the rise of personalized networking' (2001) 25(2) International Journal of Urban and Regional Research 227–251.

19 B M Compaine and D Gomery, *Who Owns the Media? Competition and Concentration in the Mass Media Industry* 3rd edn (Routledge 2000) 574; E Noam, 'Media concentration in the United States: industry trends and regulatory responses' cited in C E Baker, *Media Concentration and Democracy: Why Ownership Matters* (Cambridge University Press 2007) 28.

20 I Cram, *Liberal Democracy, Law and the Citizen Speaker: Regulating Online Speech* (Hart 2022).

21 They are accessed without charge, or relatively cheaply, through a range of portable and interconnectable devices and applications.

22 Ofcom's Online Nation 2024 Report (28 November 2024) shows that 47.4 million UK adults spend an average of 4 hours and 20 minutes online per day across smartphones, tablets and computers. 18–24-year-olds spend an average of 6 hours and 1 minute online each day. The most used services are provided by Alphabet and Meta (ch 3). Globally, it is estimated that as of February 2025 5.56 billion people accessed the internet, meaning that internet usage has more than doubled since 2010. In almost the same period the growth in social media use is even more startling: in 2010 97 million people used social networks worldwide; by February 2025 this had increased to 5.24 billion active users: Statista, 'Number of internet and social media users worldwide as of February 2025'.

creating an 'online layer through which people organise their lives … [that] influences human interaction on an individual and community level, [and] a larger societal level'.[23] Echoing the words of the Criminal Court of *New York in New York v Harris*, the 'reality of today's world' is that the internet and social media, through their platforms and applications, is 'the way people communicate'.[24] Importantly, such interactions, as Cram finds, aid in empowering non-elites in holding governmental figures to account, whilst also according them a stake in the business of government.[25]

Due, therefore, to the availability of online services, diverse non-elite voices are empowered to participate in political argument and discourse, to an unprecedented extent. Such availability has permanently altered the communication paradigm, since the ability to generate content and to communicate it to mass audiences is no longer monopolised and controlled by the traditional media.[26] Diverse ideas, opinions and content are not filtered through the mass media via traditional technology and the views of journalists and editors, but instead are rapidly developed and articulated in the exchanges of millions of people via online services.[27] These services, moreover, facilitate anonymous and pseudonymous expression.[28] Online speech has therefore played a critical role in maintaining the health of the public sphere in democratic terms by encouraging and allowing certain

23 J van Dijck, *The Culture of Connectivity: A Critical History of Social Media* (Oxford University Press 2013) 4. Penetration by social media now goes even deeper than when van Dijck was writing, as demonstrated by the recent usage figures provided by Ofcom and Statista (n 22 above).

24 *New York v Harris*, 2012 NY Misc LEXIS 1871 *3, note 3 (Crim Ct City of NY, NY County, 2012).

25 See Cram (n 20 above) 195–196.

26 Eg the United Nations Human Rights Committee has stated that the internet and social media have created 'a global network for exchanging ideas and opinions that does not necessarily rely on the traditional mass media intermediaries': Human Rights Committee, General Comment 34: Freedoms of opinion and expression, CCPR/C/GC/34 (GC 34) 12 September 2011, para 15. Similarly, the latest Ofcom UK News Consumption Report says that of the 96% of UK adults who consume news in some form, 71% use online sources, meaning that for the first time online news consumption has surpassed television (which had fallen from 75% in 2023 to 70% by 2024). Social media is a significant component of online news consumption, with 52% of UK adults using it as a news source. See Ofcom, 'News consumption in the UK: 2024 – Research Findings' (10 September 2024) 3–5.

27 See generally Koltay (n 18 above); R L Weaver, *From Gutenberg to the Internet: Free Speech, Advancing Technology, and the Implications for Democracy* 2nd edn (Carolina Academic Press 2019).

28 P Coe, 'Anonymity and pseudonymity: free speech's problem children' (2018) 22(2) Media and Arts Law Review 173–200; E Barendt, *Anonymous Speech* (Hart P2016).

actors – who possibly would not speak without the mask of anonymity or pseudonymity to protect them – to make important contributions to public discourse, and to challenge orthodox views more readily.[29] The combination of these factors has supported participatory democracy, expanding and democratising the public sphere by creating greater opportunities for individuals to engage in public discourse,[30] and for a wider variety of commercial, non-commercial and voluntary non-commodified actors to create and offer a greater diversity of content to the public.[31] In certain respects, therefore, an extremely rich speech environment has been created, very clearly in tune with furthering democratic aims, as the United States (US) Supreme Court has pointed out. In echoing Justice Stevens' judgment in *Reno v American Civil Liberties Union*[32] that the internet can enable anyone to become a 'town crier with a voice that resonates further than it would from a soap box',[33] Justice Kennedy, in *Packingham v North Carolina*,[34] found that the internet is the 'modern public square', 'one of the most important places to exchange views is cyberspace, particularly social media' and 'websites can provide perhaps the most powerful mechanisms available to private citizens to make their voices heard'.[35]

In a similar vein, the importance of online expression to the health of the public sphere, and the democratic process generally, has been emphasised by the Council of Europe's Committee of Ministers, stating that '[c]itizens' communication and interaction in online environments and their participation in activities … involving … matters of public interest can bring positive, real-life, social change'.[36] The value to

29 Eg Barendt (n 28 above) 64, 129; Human Rights Council, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, 22 May 2015, A/HRC/29/32, paras 23 and 31; E Stein, 'Queers anonymous: lesbians, gay men, free speech and cyberspace' (2003) 38 Harvard Civil Rights – Civil Liberties Law Review 159–214, 199–205; P Bernal, *The Internet, Warts and All: Free Speech, Privacy and Truth* (Cambridge University Press 2018) 220–225.

30 Cram (n 20 above) ch 4.

31 M Barnidge et al, 'Social media as a sphere for "risky" political expression: a 20-country multi-level comparative analysis' (2018) 23(2) International Journal of Press/Politics 161–182; L Bode, 'Political news in the news feed: learning politics from social media' (2016) 19(1) Mass Communication and Society 24–48.

32 *Reno v American Civil Liberties Union* (1997) 521 US 844.

33 Ibid 870.

34 *Packingham v North Carolina* 582 US_2017.

35 Ibid 1737.

36 The Council of Europe's Committee of Ministers, 'Declaration by the Committee of Ministers on the protection of freedom of expression and information and freedom of assembly and association with regard to Internet domain names and name strings' (Adopted by the Committee of Ministers on 21 September 2011) para 3.

democracy of such online participation by citizens has been accorded a particular emphasis by Cram: in rejecting the stance of 'deliberative democrats',[37] in so far as it limits the participation of non-elite groups in political discourse,[38] he contends that online communications provide disaffected individuals and organisations, excluded from elitist discourse, with a method of bringing their grievances to the attention of like-minded others.[39] A number of writers agree on the value of the contribution to participatory democracy of enabling online expression, but the notion that *therefore* top-down intervention via regulatory interference with the tech companies' market freedom should be resisted, is far from universal.[40]

Online harms inimical to the healthy functioning of democracies

But the liberation of free speech due to the availability of online platforms, partly due to the removal of the guardrails typically associated with offline forms of mass communication, also means that online services have become fertile grounds for breeding anti-democratic cyber-harms that can grow and spread rapidly.[41] The tech companies running services, such as Google, Facebook, Instagram and X, have not only proved to be far from adequate in tackling such cyber-harms, but have in some cases contributed to their proliferation, either deliberately, as can currently be said of Elon Musk's X,[42] or recklessly, regardless of their adverse impact on the health of democracies.[43] Alongside the introduction of regulatory schemes, including the OSA, hostility between the companies and various democracies has

37 Cram (n 20 above).
38 Ibid ch 2.
39 Ibid ch 4.
40 See eg Tambini (n 13 above) ch 16.
41 See, eg Abrusci (n 13 above) 3, 7, 9; P Coe, 'Tackling online false information in the United Kingdom: the Online Safety Act 2023 and its disconnection from free speech law and theory' (2023) 15(2) Journal of Media Law 213–242, 227; Dame Sara Khan, 'Societal threats and declining democratic resilience: the new extremism landscape' (*Crest Insights* 9 December 2024) 32–33 and ch 5.
42 Eg Musk's use of X to influence the outcome of the 2024 US General Election is discussed below (see n 95 and accompanying text). More recently, in January 2025, Musk used the platform to make unsubstantiated and destabilising claims that UK Prime Minister Keir Starmer and other senior politicians were 'complicit' in the 'grooming gangs' scandal: N Keate, 'UK's Starmer slams Musk's "lies" on grooming gangs' (*Politico* 6 January 2025).
43 See eg Dawes (n 1 above); Schmidt (n 12 above) 203; P Coe, 'The public sphere and the regulation of online harms and hate speech: have we opened Pandora's box?' (2022) 14(1) Journal of Media Law 50–75.

resulted.[44] Thus, accompanying the lauding of the pro-democratic benefits deriving from the availability of online services, serious concerns have arisen in the UK and globally – contrary to Cram's thesis – as to the dangers to the healthy functioning of democracies presented by some online content.[45] Providers have a clear 'capacity to influence and steer audiences', often stealthily via algorithmic choices.[46] Online content can be, as discussed, beneficial in pro-democratic terms since it is largely unmediated, lacking in editorial oversight, and often disseminated anonymously or pseudonymously.[47] But equally those very qualities, combined with online practice, can enable the proliferation of an array of cyber-harms *inimical* to democracy, including in particular the spread of mis- and dis-information,[48] while the debasing, vulgarising and polarising of political debate constitutes a key algorithmic feature.[49] This article, while acknowledging the

---

44 Eg Australia passed legislation to force Facebook to pay for news items; as a result Facebook decided in 2021 to ban news content in Australia for a period, although it then reversed its decision; but recently it is again reconsidering this ban: M Broersma, 'Meta considers Facebook news ban in Australia' (*Silicon* 1 July 2024). In the US the CEO of TikTok was questioned in front of the House Energy and Commerce Committee: D Kerr, 'Lawmakers grilled TikTok CEO Chew for 5 hours in a high-stakes hearing about the app' (*NPR* 23 March 2023). In the UK Mark Zuckerberg refused more than once to appear before Parliament in relation to the effects of fake news on UK democracy: A Hern and D Sabbagh, 'Zuckerberg's refusal to testify before UK MPs "absolutely astonishing"' *The Guardian* (London 27 March 2018).

45 J Rowbottom, 'Transposing public service media obligations to dominant platforms' in Tambini and Moore (n 13 above) ch 12, 246.

46 U Kohl, 'Toxic recommender algorithms: immunities, liabilities and the regulated self-regulation of the Digital Services Act and the Online Safety Act' (2024) 16(2) Journal of Media Law 301–335, 305–307; G Magarian, 'The internet and social media' in A Stone and F Schauer (eds), *The Oxford Handbook of Freedom of Speech* (Oxford University Press 2021) ch 19, 353.

47 It could be found that self-regulatory content moderation by the companies represents quasi-editorial functions, but the oversight is clearly far more light-touch than it is in relation to the traditional media. See S Law, 'Effective enforcement of the Online Safety Act and Digital Services Act: unpacking the compliance and enforcement regimes of the UK and EU's online safety legislation' (2024) 16(2) Journal of Media Law 263–300, 269–270; E Douek, 'Content moderation as systems thinking' (2022) 136 Harvard Law Review 526–607, 537–538.

48 That is partly because such content is conveyed by technology that is both pervasive and invasive: see P Coe, *Media Freedom in the Age of Citizen Journalism* (Edward Elgar 2021) 148.

49 Eg see Abrusci (n 13 above) 1–6; E F Judge and A M Korhani, 'Disinformation, digital information equality and electoral integrity' (2020) 19(2) Election Law Journal 240–261; S Morgan, 'Fake news, disinformation, manipulation and online tactics to undermine democracy' (2018) 3(1) Journal of Cyber Policy 39–43.

potential for regulatory over-reach, takes the view that free speech on social media can be legitimately curtailed in order to protect the health of the democracy, since the value of participatory democracy[50] is undermined when those participating do so on the basis of false information fed into the political process[51] or when politicians respond defensively (possibly by voting against their principles) to abusive and threatening forms of participation. Threats against politicians have further been found to foster increasing political polarisation and to present 'a threat to democratic stability and to democracy at large, impacting trust, engagement and participation'.[52] Democratic health is also damaged when, in response to terrorism or hate speech, it defensively abandons democratic ideals. This stance establishes a benchmark against which to assess the OSA's success or failure in tackling the three specific anti-democratic harms discussed below; the OSA's attempt to tackle them is evaluated in the second section of this article. The following discussion explores those harms, indicating that reliance on self-regulation alone has failed to curb them or has even accorded them encouragement. The three harms explored below have been selected because they are, as indicated, of anti-democratic tendency, and have, consequently, been identified globally as requiring a remedy;[53] in the UK they were relied on as a key justification for the introduction of the OSA.[54] They are also harms – aside from terrorism and hate speech, which were already the subject of a web of offences – that were singled out as requiring the introduction of new criminal offences in the OSA itself, as discussed in the second section, below.

Targeting politicians: cyber-bullying, online abuse, deep-fake pornography

The use of social media to direct abusive communications at individuals, often due to their status, can be identified as a particular problem in relation to UK politicians, partly due to their impact on democracy. Such user-generated content has included death or rape

50 See further on that value J Habermas, *Philosophical Introductions: Five Approaches to Communicative Reason* (Polity Press 2018) 109–110; A Kenyon, *Democracy of Expression* (Cambridge University Press 2021) ch 3.

51 See n 90 below and associated text.

52 See 'Violence against politicians', Council of Europe 20–21 March 25.

53 Magarian (n 46 above) ch 19.

54 See nn 4 and 5 above and accompanying text; see also *White Paper: Full Government Response to the Consultation* (n 3 above) 'Joint ministerial foreword', 3.

threats, sometimes becoming a feature of their everyday life.[55] The very immediacy and ease of use of the medium means that individuals are deploying platforms to give vent to unfiltered anger directed at MPs as high-profile figures making controversial decisions,[56] often in order to intimidate and seek to silence them.[57] Communicating via social media, especially anonymously, clearly fosters disinhibition and an unaccountability unavailable offline,[58] an increasing problem currently in relation to the online targeting of MPs.[59]

55 See eg R Southern and E Harmer, 'Twitter, incivility and "everyday" gendered othering: an analysis of tweets sent to UK Members of Parliament' (2021) 39(2) Social Science Computer Review 259–275. See further the subsection 'Online threats and abuse targeting politicians' below. Cyber-bullying or 'cyber-stalking' has been found in a range of studies to constitute a serious problem in terms of the psychological harm suffered by victims; see House of Lords (n 14 above) paras 246–249; 'The Impact of Online Abuse: Hearing the Victims' Voice' (Victims Commissioner 30 May 2022). It was one of the harms listed by the *White Paper* as requiring regulation, see *White Paper* (n 3 above) eg para 1.15, 16, para 7.37, 73, paras 7.43–7.47, 75.

56 See eg E Esposito and R Breeze, 'Gender and politics in a digitalised world: Investigating online hostility against UK female MPs' (2022) 33(3) Discourse and Society 303–323.

57 Eg individuals might combine to target an individual on multiple occasions (mass trolling or raiding), or false Facebook or X accounts might be set up for that purpose: L H Sun and P Fichman, 'The collective trolling lifecycle' (2020) 71(7) Journal of the Association for Information Science and Technology 727–868.

58 J Suler, 'The online disinhibition effect' (2004) 7(3) CyberPsychology and Behaviour 321–326, 322; S Levmore, 'The anonymity tool' (1996) 144 University of Pennsylvania Law Review 2191–2236; S Levmore, 'The internet's anonymity problem' in S Levmore and M Nussbaum (eds), *The Offensive Internet* (Harvard University Press 2010) 50, 54–55; D K Citron, *Hate Crimes in Cyberspace* (Harvard University Press 2014) 4–12.

59 See eg A Dickson 'UK wrestles with online anonymity in wake of MP's killing' (*Politico* 19 October 2021); N Johnston and N Davies, 'Intimidation of candidates and voters' (House of Commons Library Research Briefing 8 April 2024) 7–9.

Of particular significance is the compelling evidence that female MPs are being disproportionately targeted.[60] It has been found that 'today's toxic virtual environment poses a real risk to the future of women in politics'[61] and to democracy more generally.[62] Certain female MPs have also recently been found to be the latest victims of deep-fake pornography.[63] There is evidence that these forms of online abuse targeting female politicians in particular are deterring women from entering politics and are a factor in coercing them to leave.[64] A number of activist organisations have further pointed out that different characteristics, such as being both a woman and from an ethnic minority, can intersect to mean that the experience of online abuse creates greater psychological harm, again a matter disproportionately affecting female politicians.[65] The scale of online abuse of politicians in general, especially on X, can readily be viewed as having an anti-democratic impact, particularly when it deters them from speaking on certain controversial subjects or places pressure on them to vote against their principles.[66]

---

60 See: Demos, 'The scale of online misogyny' (26 May 2016); S Laville, 'Research reveals huge scale of social media misogyny' *The Guardian* (London 26 May 2016); Oral Evidence to the Home Affairs Select Committee, 'Hate crime and its violent consequences' (2017) HC 609. Executives from Google, Facebook and Twitter gave oral evidence, covering online abuse directed in particular at female MPs. See also reports relating to female MPs, referencing the impact of online abuse on them in 2019: M Nadim and A Fladmoe, 'Silencing women? Gender and online harassment' (2021) 39(2) Social Science Computer Review 245–258. Figures in the first national analysis of the scale of violence against women and girls by the National Police Chiefs' Council, released on 16 July 2024, highlighted the problem of young men being 'radicalised' online by influencers such as Andrew Tate: see V Dodd, 'Violence against women a national emergency police say' *The Guardian* (London 23 July 2024.

61 C Julios 'Ignoring online abuse of women MPs has dire consequences' (*LSE Blog* 17 May 2023). See also Ofcom, 'A safer life online for women and girls: practical guidance for tech companies' (25 February 2025) para 2.21, 13.

62 Fawcett Society, A House for Everyone: A Case for Modernising Parliament (January 2023) ch 5.

63 Ofcom (n 61 above); C Newman, 'Top female politicians victims of deep fake porn' (*Channel 4 News* 1 July 2024). Victims include: the Labour Deputy Leader, Angela Rayner; the former Education Secretary, Gillian Keegan; the former Commons Leader, Penny Mordaunt; the former Home Secretary, Priti Patel. The report noted that many of the images had been online for several years and had attracted hundreds of thousands of views.

64 Julios (n 61 above).

65 Esposito and Breeze (n 56 above); Johnston and Davies (n 59 above) 7–9; Fawcett Society (n 62 above) ch 5. In the context of dis-information specifically, see Abrusci (n 13 above) 9–10.

66 See Khan (n 41 above) 50; P Lynch, P Sherlock and P Bradshaw, 'Scale of abuse of politicians on Twitter revealed' (*BBC News* 9 November 2022).

Anti-democratic impacts of fostering terrorism and promulgating hate speech online

Terrorist propaganda,[67] live-streaming of videos of terrorist attacks[68] and promulgation of hate speech[69] can clearly reach far wider audiences online than offline. The failures of self-regulation are perhaps most clearly demonstrated by highlighting the widespread and increasing proliferation of such content. The Organisation for Security and Co-operation in Europe (OSCE) found in 2024 that 'the Internet has become one of the main tools in the arsenal of contemporary terrorist organizations'[70] and envisaged a clear likelihood of acceleration of online activity among terrorist actors in future. The UK Government claimed in the *White Paper* that all five domestic terrorist incidents in 2017 had online elements, including online radicalisation by international groups, such as ISIS.[71] The *White Paper* also pointed out that 'the terrorist group Daesh used over 100 platforms in 2018, making use of a wider range of more permissive and smaller platforms'.[72]

The role of terrorism in destabilising democracies and in leading to 'defensive democracy' is the subject of an extensive literature,[73] so the

---

67 The House of Lords Communications and Digital Committee inquiry into Freedom of Expression Online, written evidence (FEO0012, 8 January 2021) noted (para 10(c)) that Twitter actioned 95,887 unique accounts related to the promotion of terrorism/violent extremism between January and June 2019: Rule Enforcement, Twitter, August 2020.

68 The Christchurch terrorist attack was carried out by a far-right extremist, killing 51 people in two mosques in Christchurch in New Zealand; the terrorist live-streamed the attack on Facebook: G Macklin 'The Christchurch attacks: livestream terror in the viral video age' (2019) 12(6) Combating Terrorist Center 18–29.

69 The design of the systems used by the tech companies has been found to amplify abusive communications, including hate speech, and promote their dissemination. As found in evidence presented to the House of Lords Select Committee on the OSB: 'Algorithms distribute, amplify and suppress the visibility of content in opaque ways, meaning that users often have limited control over what they see'; 'content curation algorithms are designed to engage, and it turns out that the most engaging content is really toxic content': House of Lords (n 14 above) para 226 – citing P-J Ombelet, 'The chilling effects of content policing by social media' (KU Leuven CiTiP 5 July 2016) – and para 227 – citing evidence from Dr Carissa Véliz, Associate Professor in Philosophy, University of Oxford.

70 In OSCE Secretariat, 'Countering the use of the internet for terrorist purposes' (2024). See also Khan (n 41 above) ch 5.

71 It also stated that 'online terrorist content remains a feature of contemporary radicalisation': *White Paper* (n 3 above) para 1.9.

72 Ibid para 1.10.

73 H Fenwick 'Terrorism and the control orders/TPIMs saga: a vindication of the Human Rights Act or a manifestation of "defensive democracy"?' (2017) 4 Public Law 609–626; C Walker 'Keeping control of terrorists without losing control of constitutionalism' (2007) 59(5) Stanford Law Review 1395–1463.

threat it poses to democracy need not be rehearsed further here, but the recent direct terrorist threat to MPs, fostered by online activity, is obviously anti-democratic in tendency and is less well documented. The killing of MP David Amess in his constituency surgery by Ali Harbi Ali, an extremist Muslim reportedly radicalised online, aided in bringing the topic of online threats to MPs and the use of the internet to groom recruits and foster radicalisation[74] – exacerbated during Covid-19 lockdowns[75] – to parliamentary and public attention. Recently, Mike Freer resigned as an MP due to multiple, largely online, threats made against him, citing his narrow escape from Harbi Ali.[76] Amess's murder and Freer's experience follow a long line of the targeting of MPs by terrorists,[77] often online, and recently the threat has risen in relation to the stance, or perceived stance, of MPs relating to the situation in Gaza, meaning that political debate on that topic may be inhibited,[78] as was canvassing in the UK 2024 general election.[79]

The anti-democratic impact of hate speech is also quite readily apparent: a range of studies have found links between the silencing and withdrawal from online arenas of debate, in particular political ones, of various groups[80] due to the impact of online hate speech,[81] and, further,

74 HM Prison and Probation Service, 'Exploring the role of the internet in radicalisation and offending of convicted extremists' (Ministry of Justice Analytical Series 2021). It found, for example, at paras 4.1 and 5.1, that 'the role of the internet was increasingly prominent in the radicalisation of convicted extremists in England and Wales'. Amess was killed on 15 October 2021.

75 See comment to this effect by Security Minister Damian Hinds to *Sky News*, reported by K Feehan, 'More people have been radicalising themselves online at home during Covid lockdowns, Security Minister Damian Hinds says' (*MailOnline* 16 November 2021).

76 In an interview with *GB News*, Mr Freer called for social media firms to take more action against content that incited violence against MPs: see C Geiger, 'Hate faced by MP Mike Freer is attack on democracy, says Downing Street' (*BBC News* 1 February 2024); 'Editorial: The Guardian view on threats to MPs: Mike Freer's experience should serve as a warning' *The Guardian* (London 4 February 2024).

77 'Editorial' (n 76 above). This provides several examples, including of Andy Pennington and Nigel Jones (2000), Stephen Timms (2010), Jo Cox (2016) and Rosie Cooper (2022). See also Khan (n 41 above) 17, 50, 64.

78 E Courea and J Halliday, 'British MPs fearful of violent attacks as tensions over Gaza war increase threats' *The Guardian* (London 23 February 2024.

79 K Stacey, 'Labour condemns harassment of its candidates in pro-Palestinian areas' *The Guardian* (London 3 July 2024).

80 C Carlson 'Hate speech as a structural phenomenon' (2020) 54 First Amendment Studies 217–224, 217.

81 A Siegel 'Online hate speech' in N Persily and J Tucker (eds), *Social Media and Democracy* (Cambridge University Press 2020) ch 4; A Brown, 'What is so special about online (as compared to off-line) hate speech?' (2018) 18(3) Ethnicities 297–326, 304.

that journalists, politicians and bloggers have been disproportionately targeted.[82] A recent dramatic rise in the volume of hate speech online has been identified by several commentators;[83] it was implicated in the 2024 riots and ensuing prosecutions in the UK.[84] A recent report presented to the United Nations Human Rights Council commented on that rise and found: 'In many countries, three quarters or more of the victims of online hate speech are members of minority groups. Women belonging to these groups are disproportionately targeted.'[85] The Council of Europe's 2024 Recommendation on Combating Hate Crime[86] emphasised the responsibility of online services to remove content falling within the scope of hate crimes;[87] that plea has also been reiterated by the Chief Executive of Ofcom.[88] Despite such pleas, self-regulation appears to have been ineffective in curbing online hate speech, and the same can be said of the voluntary European Union (EU) Code of Conduct on Hate Speech.[89]

---

82 T Isbister et al, 'Monitoring targeted hate in online environment' (March 2018); Johnston and Davies (n 59 above) 7–9.

83 Khan (n 41 above) 31; Alan Turing Institute, 'Detecting East Asian prejudice on social media' (nd); K Müller and C Schwarz, 'Fanning the flames of hate: social media and hate crime' (2021) 19(4) Journal of the European Economic Association 2131–2167; A Harel, 'Hate speech' in A Stone and F Schauer, *The Oxford Handbook of Freedom of Speech* (Oxford University Press 2021) 455–476; *White Paper: Full Government Response to the Consultation* (n 3 above) paras 2.3, 2.29.

84 R Fern, 'Riots in the UK: online propagandists know how to work their audiences: this is what we are missing' (*Inforrm* 5 August 2024); House of Commons Library, 'Policing response to the 2024 summer riots' (UK Parliament 9 September 2024). See also n 200 below.

85 'Hate speech, social media and minorities' (2021) Report by Special Rapporteur Fernand De Varennes presented to the outcomes of the 13th Forum on Minority Issues.

86 Recommendation CM/Rec(2024)4 of the Committee of Ministers to Member States on Combating Hate Crime, 7 May 2024.

87 Ibid para 68. See also para 69 as to duties to remove the content in question.

88 Prior to the enactment of the OSA, Dame Melanie Dawes – the Chief Executive of Ofcom – said that the proliferation of hate speech on social media platforms brought the 'need for regulation' into 'sharper focus' and that platforms must do more to combat such incidences (n 1 above).

89 T Quintel and C Ullrich, 'Self-regulation of fundamental rights? The EU Code of Conduct on Hate Speech, related initiatives and beyond' in B Petkova and T Ojanen, *Fundamental Rights Protection Online: The Future Regulation of Intermediaries* (Edward Elgar 2020) 197–229.

The adverse impact of false information on democratic processes[90]

Although false information has been prevalent for centuries,[91] today the internet provides the ideal environment for it to grow and spread quickly online and offline.[92] The volume of online false information generated about Covid-19 was one of the drivers behind the initial inclusion of false information 'that could cause significant harm to an individual' within the scope of the UK online harms regime, as originally conceived.[93] Many other high-profile examples of the anti-democratic impact of mis- and dis-information arise. Cambridge Analytica, for instance, harvested over 50 million user profiles without Facebook's permission, and manufactured sex scandals and dis-information to influence voters in elections globally, including in the UK.[94] The dis-information-for-profit market is extensive, in which private contractors, employed by companies and politicians, have used social media to manipulate elections worldwide. The practice was predicted – as it transpired, accurately – to escalate during the UK and US 2024 general elections.[95] The damage this is doing to the public sphere and democracy in general was highlighted by the 2024 Oxford

90 'False information' is a generic term that is applied to either or both dis-information and mis-information, and, less commonly, 'mal-information'; see Council of Europe, 'Dealing with propaganda, misinformation and fake news' (nd).

91 P Zanker, *The Power of Images in the Age of Augustus* (University of Michigan Press 1990); L Becker, *The Role of Propaganda in Ancient Empires: Influencing Power and Public Perception* (Ancient History Guide 2024); Bernal (n 29 above) ch 9.

92 P Bernal, 'Fakebook: why Facebook makes the fake news problem inevitable' (2018) 69(4) Northern Ireland Legal Quarterly, 513–530, 516–519; A Guess and B Lyons, 'Misinformation, disinformation, and online propaganda' in N Persily and J Tucker (eds), *Social Media and Democracy* (Cambridge University Press 2020) ch 2; Khan (n 41 above) 32–33 and ch 5.

93 It originally fell within the OSB's 'legal but harmful' provisions: see *White Paper: Full Government Response to the Consultation* (n 3 above) paras 34, 84–85, 34, 84–85; N Dorries, Statement UIN HCWS19. For a theoretical discussion on the regulation of 'legal but harmful' content, see K Konstantinos, 'Online harm, free speech, and the "legal but harmful" debate: an interest-based approach' (2024) 16(2) Journal of Media Law 390–416.

94 P N Howard, *Lie Machines* (Yale University Press 2020) 12.

95 Eg see T Graham, 'Elon Musk's flood of US election tweets may look chaotic. My data reveals an alarming strategy' (*The Conversation* 6 November 2024); D Milmo and A Hern, 'Elections in UK and US at risk from AI-driven disinformation, say experts' *The Guardian* (London 20 May 2023). See also House of Commons Digital, Culture Media and Sport Committee, *Disinformation and 'Fake News': Final Report*, HC 1791, 18 February 2019, 68–77. For a global perspective on this issue, see S Bradshaw and P N Howard, The Global Disinformation Order 2019: Global Inventory of Organised Social Media Manipulation (Oxford Internet Institute and University of Oxford 2019).

University *Reuters Institute Digital News Report*, which found that 59 per cent of the 95,000 people surveyed worldwide were concerned as to differentiating between real and false news online.[96] Thus, while, as discussed above, the internet has, in some respects, contributed to the democratisation of the public sphere, not least in supporting rights to receive information,[97] its role in proliferating false information can also be viewed as having a destabilising impact on democracy by heightening ethnic and nationalistic tensions, weakening public trust in journalism, in democratic institutions and in electoral outcomes.[98]

It is apparent even from this brief discussion that self-regulation implemented by the platforms themselves has failed to address the problem of the promulgation of false information online.[99] Equally, voluntary, non-binding co-regulatory codes of conduct on this matter, including the European Commission's Code of Practice on Disinformation,[100] have also proved to be largely ineffective in practice. For instance, the European Commission has criticised the

96 N Newman et al, Reuters Institute Digital News Report (Reuters Institute/University Oxford 2024). The report (at 17–18) finds that concerns about how to distinguish between trustworthy and untrustworthy content in online platforms are highest for TikTok and X when compared with other online networks. Specifically, among those surveyed, there was particular concern about differentiating between real and false information relating to politics. Similarly, a recent Ipsos and UNESCO survey conducted in 16 countries found that 87% of people were concerned about the impact of dis-information on elections in their country, with 89% urging their government and regulators to improve trust and safety on social media platforms during elections. See 'Survey on the Impact of Online Disinformation and Hate Speech' (Ipsos/UNESCO September 2023).

97 Certain tech companies are under a statutory requirement to promote 'content of democratic importance' (see s 17 OSA), which would include facilitating the provision of information relating to the democratic process. While the companies as private actors are not bound by art 10 of the European Convention on Human Rights, covering rights to receive information, s 17 indicates that they have a role in supporting the realisation of art 10 rights, with which the legislation itself is compatible (s 3 Human Rights Act 1998; see also s 19 HRA).

98 See Howard (n 94 above) 18; Abrusci (n 13 above) 1–10; Guess and Lyons (n 92 above); former Facebook employee Sophie Zhang testified about this problem to Parliament in October 2021: see Joint Committee on the Draft Online Safety Bill, HL Paper 129, HC 609, 14 December 2021, para 105.

99 Eg the Oxford Technology and Elections Commission found that many of the self-regulatory measures taken by social media platforms have failed to prevent the spread of dis-information: S Hoffmann et al, 'The market of disinformation' (OxTec October 2019); E Shattock, 'Self-regulation 2:0? A critical reflection of the European fight against disinformation' (*Harvard Kennedy School Misinformation Review* 31 May 2021).

100 This Code was updated in 2022 in line with the European Union's DSA. This Act came into force on 25 August 2023 for very large online platforms, such as X and Facebook. It became fully applicable to other entities on 17 February 2024. The UK is not subject to it due to its exit from the EU.

inconsistent and incomplete application of the required commitments in the Disinformation Code from the services.[101] It appears that online services have used the code and self-regulation reflecting it for self-serving interests, including enhancing or preserving reputations and to avoid more direct and onerous regulatory oversight.[102] The responses of the tech companies to this problem clearly vary, but currently the volume of false information online is likely to rise. Musk's X appears to be rejecting the notion of deploying self-regulation to curb the promulgation of false information online;[103] Meta is following suit in

101 See European Commission, Assessment of the Code of Practice on Disinformation – Achievements and Areas for Further Improvement (Commission Staff Working Document SWD(2020) 180 final) 7–19; P Cavaliere, 'The truth in fake news: how disinformation laws are reframing the concepts of truth and accuracy on digital platforms' (2022) 3(4) European Convention on Human Rights Law Review 481–523, section 3.

102 E Shattock, 'Fake news in Strasbourg: electoral disinformation and freedom of expression in the European Court of Human Rights' (2022) 13(1) European Journal of Law and Technology 3.

103 In 2021 Twitter (as it was then known) introduced Birdwatch, a fact-checking model rebranded on X as Community Notes, which delegates fact-checking to the 'community' through approved contributors who identify content deemed to be false or misleading by attaching notes providing more context (see 'About Community Notes on X'). Community Notes became widespread in 2023, after Musk purchased the platform. The model's anti-democratic potential has been highlighted by the Center for Countering Digital Hate, in its October 2024 report 'How X's Community Notes system falls short on misleading election claims'. In December 2023 the European Commission opened ongoing formal proceedings to assess whether X's use of Community Notes may have breached the DSA (European Commission Press Release, 'Commission opens formal proceedings against X under the Digital Services Act' (18 December 2023). How the model sits with the OSA is currently unclear. Due to the Act's limited scope in respect of misinformation, the key compliance question for regulated services is whether it complies with its duties regarding illegal harms and the protection of children, which can be satisfied with or without content moderation (see ss 10(2)–(3) and 12(2)–(3)). Since the Act does not prescribe required steps for compliance, and the recommendations set out in the Illegal Harms Content Codes of Practice (December 2024) are not mandatory (although Ofcom says that services that do follow the Codes will be treated as compliant with the relevant duties), the regime determines that Ofcom will assess the measures taken by a service on a case-by-case basis. The illegal content duties came into force on 17 March 2025 (see Ofcom, 'Enforcing the OSA: platforms must start tackling illegal material from today').

the US,[104] reportedly in a bid to curry favour with President Donald Trump.[105]

Inherent failures of self-regulation in addressing these harms

The discussion above indicates that self-regulation by the tech companies has rendered them the arbiters of free speech online, failing to strike an effective balance between curbing the online anti-democratic harms discussed above while also preserving online free expression. The companies clearly take varying approaches to self-regulation; some platforms remove harmful content more rapidly than others, in accordance with their codes; at the same time individual companies' policies may change rapidly.[106] This inconsistent approach has often led to a lack of restraint of online content, sometimes preserving the benefits of free speech, thereby serving democratic ends, but also enabling the promulgation of content undermining them. The efficacy of self-regulation is thus, perhaps almost inevitably, both highly questionable and dogged by inconsistency.[107] The expectation that private companies, set up in order to make a profit, and operating on the basis that online content is viewed as a marketable commodity, will voluntarily restrict online speech clearly runs counter to their

104 When announcing its new fact-checking policy (to 'dramatically reduce the amount of censorship' and 'restore free expression' on its platforms), Meta directly referenced X's Community Notes approach, albeit that it said that it has 'no immediate plans' to dispense with its third-party fact-checkers in the UK or the EU. See K Albury and J Williams, 'Meta's shift to "community notes" risks hurting online health info providers more than ever' (*The Conversation* 16 January 2025).

105 The *New York Times* reported that Meta notified Trump's team of its policy change before making the change public: M Isaac and T Schleifer, 'Meta says it will end its fact-checking program on social media posts' *New York Times* (New York 7 January 2025).

106 See Bernal (n 29 above) 127; S Zuboff, *The Age of Surveillance Capitalism* (Profile Books 2019) 48–50, 217–220. Eg on 7 January 2025 Meta made sweeping changes to its policy on Community Standards – Hateful Conduct. These changes include permitting LGBTQ+ persons to be called mentally ill, transgender people to be called "it" and women to be referred to as property in user-to-user communications on Meta's platforms: S Sharma, 'Meta's recent changes to its Hateful Conduct Community Standards place marginalised groups at serious risk and likely breach its duties under the Online Safety Act' (*Inforrm* 14 January 2025).

107 Bernal (n 29 above) 247–248. Their efficacy is also difficult to assess with accuracy.

corporate values.[108] Adherence to such values, as opposed to those rooted in the furtherance of democracy, inevitably affect the nature of the infrastructures at issue: the prevalent business model requires maximisation of user engagement and therefore of advertising revenue.[109]

Further, the services are owned by only a handful of companies;[110] as Pickard puts it: 'No firms have ever wielded so much power over our communication and information infrastructures.'[111] Control over digital media and infrastructure has been allowed to become concentrated largely in the hands of a very small number of US tech oligarchs. Self-regulation clearly cannot restrain those billionaire owners of the companies from taking decisions, regardless of their anti-democratic effects, that will enhance profitability, sometimes in response to political changes, such as the installation of Trump as US President.[112]

The responses of the companies to the OSA scheme will clearly vary; some will be much less inclined to comply than others,[113] but they are in any event highly unlikely to abandon the continuance of self-regulation.[114] The regulated services will continue to rely on their own codes, probably with modifications, in order to seek to avoid Ofcom's interventions, or in some cases to test the strength or otherwise of the

---

108 Eg R Mansell et al in their report 'Information Ecosystem and Troubled Democracy' (Observatory on Information and Democracy 3 December 2024) found that data monetisation interests are behind the way information ecosystems are operated without respect for the fundamental rights of content producers *and* the rights 'of others'. See also Bernal (n 29 above) 95–101; K Klonick, 'The new governors: the people, rules and processes governing online speech' (2018) 131 Harvard Law Review 1599–1670, 1665; Koltay (n 18 above) 180–183; Coe (n 48 above) 76–79.

109 Coe (n 48 above); Zuboff (n 106 above) ch 7; Law (n 47 above) 269–270; Kohl (n 46 above) 305–307; K Hill, *Your Face Belongs To Us* (Simon & Schuster 2023) xvi–xvii; I Katsirea, *Press Freedom and Regulation in a Digital Era: A Comparative Study* (Oxford University Press 2024) 37; M Carlson, 'Facebook in the news' (2018) 6(1) Digital Journalism 4–20, 13.

110 Eg Alphabet owns Google and YouTube; Meta owns Facebook, WhatsApp and Instagram.

111 Pickard (n 13 above) 323.

112 See nn 103–105 above. See also J Ryan, 'Big tech is picking apart European democracy, but there is a solution: switch off its algorithms' *The Guardian* (London 14 January 2025); B Montgomery, 'Why did Mark Zuckerberg end Facebook and Instagram's factchecking program?' *The Guardian* (London 7 January 2025).

113 As explained in nn 47 and 103, in respect of content moderation under the OSA regime, there is significant potential for inconsistent approaches among the tech companies.

114 The *White Paper* proposed that self-regulation would continue alongside the new regime (n 3 above) para 2.10, 35.

OSA regulatory scheme.[115] The following, second, section focuses on illegal content online to be targeted under the scheme in order to address the online anti-democratic harms outlined above. In so doing it probes flaws and gaps in the scheme, which leave open some leeway for self-regulation to continue without top-down intervention, enabling such harms to continue to manifest themselves.

THE NEW OSA REGULATORY SCHEME: EFFICACY IN BALANCING FREEDOM OF EXPRESSION WITH RESTRAINT OF ANTI-DEMOCRATIC HARMS?

Introduction

This section will identify three key weaknesses in the OSA drafting. Firstly, only the larger tech firms come within the provisions that impose the most exacting duties. Secondly, the illegal content duties imposed are not sufficiently protective in terms of combatting anti-democratic harms since the duties are dependent on finding that content is within the scope of relevant poorly drafted criminal offences; they are also subject to the weakening impact of the proportionality provisions. Thus, the section proceeds to argue, using three examples of such harmful material, that the OSA accords the tech companies too much leeway in terms of content moderation, which is likely to lead in many instances to its continued presence on the platforms. Thirdly, the section will find that the 'legal but harmful' provisions share a number of the weaknesses already identified in relation to the illegal content ones.

The categorisation scheme: flawed in pro-democratic terms

This new regulatory framework is founded on the creation of categories of service providers, determining the duties to which each one is subject. The OSA purports to cover the providers of 'internet services',[116] separating those providers into user-to-user services (services that enable a user to communicate with another user, such as X, Facebook, and Instagram)[117] and search services (such as Google).[118] However, due to the categorisation scheme, some services may fall outside scope entirely, while the duties created by the scheme do not apply equally to all categories of service. Ofcom is required to categorise these services

115 See eg M Savage, 'Tech giants told UK online safety laws "not up for negotiation"' *The Guardian* (London 11 January 2025).

116 OSA s 226.

117 Ibid s 3(1).

118 Ibid s 229. It also covers services displaying 'regulated provider pornographic content': ibid pt 5.

by establishing a register that will distinguish them as: Category 1, user-to-user services only; Category 2A, search services and user-to-user services which include a search engine; and Category 2B, user-to-user services.[119] Under the OSB, the category a service fell into was restrictively determined by user numbers *and* functionality, as well as other factors that the Secretary of State deemed relevant.[120] Due, however, to an amendment to schedule 11 and section 97(4) of the OSB, then included in the OSA,[121] Ofcom was obliged to consider functionality independently of UK user numbers when determining the categorisation of a service.[122]

Facebook, Instagram and other large mainstream platforms will be classified as Category 1 services and will therefore attract a range of further duties, going beyond the illegal content ones, as discussed at certain points below.[123] There remains, however, significant latitude in the regime in respect of categorisation, affecting the duties of some smaller services. This stems from the fact that schedule 11 requires the Secretary of State to make regulations specifying the Category 1, 2A and 2B threshold conditions,[124] to be prescribed by a post-enactment statutory instrument, which was promulgated in December 2024.[125] Such secondary legislation made outside Parliament is clearly more susceptible to pressure and lobbying from the services; it is obviously not subject to the same standard of publicly accessible,

119 Ibid s 95.

120 Ibid sch 11, para 1.

121 This amendment was tabled by Baroness Morgan of Cotes.

122 OSA, s 97(4): 'If the regulations under paragraph 1(1) of Schedule 11 specify that a service meets the Category 1 threshold conditions if any one condition about number of users or functionality is met (as mentioned in paragraph 1(4)(a) of that Schedule)—(a) subsection (2) applies as if paragraph (b) were omitted, and (b) subsections (3) and (8) apply as if the reference to the conditions in subsection (2) were to the condition in subsection (2)(a)'. However, under the regulations promulgated, functionality is relevant alongside user numbers. See n 126 below.

123 Under s 7(5): 'All providers of Category 1 services must comply with the following duties in relation to each such service which they provide—the duty about illegal content risk assessments set out in section 10(9) … (e) the duties to protect content of democratic importance set out in section 17, (f) the duties to protect news publisher content set out in section 18, (g) the duties to protect journalistic content set out in section 19, (h) the duties about freedom of expression and privacy set out in section 22(4), (6) and (7). See n 125 below as to the regulations now made determining categorisation.

124 OSA, sch 11, para 1 and s 224(3). On 25 March 2024 Ofcom stated that it 'submitted its research and advice on categorisation to the Secretary of State on 29 February 2024. The Secretary of State will consider this advice before setting regulations for categorisation.' Now see n 125

125 OSA (Category 1, Category 2A and Category 2B Threshold Conditions) Regulations 2025. The draft SI 2025 was laid on 16 December 2024 under OSA, s 225(8).

parliamentary scrutiny as primary legislation. The result, it appears, from the statutory instrument, is that some small, but high-harm, platforms may *not* be categorised as Category 1 services,[126] or may fall entirely outside scope.[127] That raises a number of concerns, including as to the ability of content of anti-democratic tendency to flourish on some smaller services which, if outside scope, would continue to rely on self-regulation alone. Indeed, this already seems to be becoming a reality. In February 2025 the Government's Regulatory Policy Committee issued an opinion, observing that the thresholds mean that fewer platforms are now expected to fall into Categories 1 and 2A. Significantly, the Act's tiered approach to categorisation could encourage potentially harmful sites or platforms to choose to remain small, or to disaggregate into many smaller sites, to avoid having to comply at all.[128] Moreover, along with other limitations on their duties thereby created, they will not be subject to the duty to promote 'content of democratic importance' if falling outside Category 1,[129] regardless of whether they are within scope. If that duty was intended to aid in creating a balance between combatting online harms and protecting

126 The SI (ibid) made under OSA, s 97(4), indicates that in relation to certain user-to-user services this is the case. Under reg 3(1)(a),(b), Category 1 services are those that have an average number of monthly active UK users exceeding 34 million and use a content moderator system or have an average number of monthly active UK users exceeding 7 million, use a content recommender system and provide a functionality for users to forward or share regulated user-generated content on the service with other users of that service. The threshold conditions for Category 2B services are prescribed in reg 5(a), (b), and are met where, in respect of the user-to-user part of that service, it has an average number of monthly active UK users exceeding 3 million, and provides a functionality for users to send direct messages to other users of the same service which is designed so that messages cannot be encountered by any other users of that service unless further action is taken by the user who sent the message or a user who received the message. Thus, less popular or decentralised platforms, such as Rumble, Mastadon, Discord, Reddit and Tumblr, may fall within Category 2B or may fall entirely outside scope once Ofcom has compiled the data on UK user numbers.

127 Eg Threads, owned by Meta, is not, it appears, currently within scope; in November 2024 it had an average number of monthly active UK users that was under 3 million. This may also invite comparisons with broader frameworks elsewhere – eg Australia's Online Safety Act 2021, in relation to children, brings a wide range of services within scope: it already covered social media but now under the 2021 Act it covers other online services. Reddit, Facebook, Instagram, WhatsApp, OnlyFans, Bumble and even Zoom or Microsoft Teams are all covered by the new Australian requirements. See eg G Fraser, 'X refused to take down video viewed by Southport killer' (*BBC News* 24 January 2025).

128 From a previous estimate of 20 down to 15, although with more platforms in Category 2B (up from 15 to 28). See Regulatory Policy Committee Opinion on the Department for Science, Innovation and Technology's Impact Assessment (IA) in Respect of the Regulations, 19 February 2025.

129 See OSA, ss 17(2), 19, 22.

free expression where it is defined as of democratic importance,[130] it can clearly have no role in so doing in relation to certain smaller services. Non-prioritisation of that duty is therefore apparent. Clearly, the contribution of the categorisation scheme to the balance struck by the OSA between protection from online harms and promotion of free expression, especially of democratic importance, is questionable.

Transforming the relationship between online services and expression-based offences: the new duties relating to 'illegal content'

Until the OSA came into force, the response to online expression deemed harmful remained largely divided between a criminal law[131] and a self-regulation-based one. In contrast to the position of traditional media bodies as publishers,[132] expression-based offences had no direct application to the hosting online services themselves, as opposed to individual posters. Therefore, as far as the services were concerned, user-generated hate speech, threatening expression, false information or terrorist content, regardless of the anti-democratic impacts created, could flourish online, subject only to self-regulation. The traditional divide, long recognised in free speech theory between distributing content and publishing it,[133] implicitly benefited online

130 The definition is also itself of narrow scope since it is confined to debate in the UK: under s 17 'content is "content of democratic importance", in relation to a user-to-user service, if—(a) the content is—(i) news publisher content in relation to that service, or (ii) regulated user-generated content in relation to that service; and (b) the content is or appears to be specifically intended to contribute to democratic political debate in the United Kingdom or a part or area of the United Kingdom'.

131 Thus, the creators or publishers of online content fall within the criminal liability attracted currently by some online expression, but criminalisation generally does not cover the online platforms and services enabling access to it.

132 Such offences are reflected in their regulatory codes. Broadcasters are regulated by Ofcom; the press is subject to self-regulation. There are two press regulators: one is the Independent Press Standards Organisation; the Press Recognition Panel has approved the other regulator, Impress. For further commentary, see J Rowbottom, *Media Law* 2nd edn (Hart 2024) ch 7; P Wragg, *A Free and Regulated Press: Defending Coercive Independent Press Regulation* (Hart 2020).

133 Some distinctions, however, were made in UK criminal law between the publication and creation of content via the *mens rea* elements of certain speech-based offences. For example, s 1 of the Terrorism Act 2006 creates the offence of encouragement of terrorism, but the publisher, as opposed to the creator of the content, may fall outside the scope of the offence due to lack of the requisite *mens rea*. On the other hand, under eg s 1 of the Obscene Publications Act 1959 both the publisher of the content and the creator who possesses it for gain are within scope. The position is similar under the Contempt of Court Act 1981, s 2(1): the material must be published although it is irrelevant whether that is for gain or not.

services since as hosts they were placed in effect in the position of most distributors.

But this new regulatory regime now partially breaks down that traditional distinction between distributing/hosting and creating/publishing material, partly by creating the concept of 'illegal content', although, as the *White Paper* noted, 'publisher' levels of liability do not apply.[134] The new OSA duties of care imposed on certain services to protect users from such content are, if taken at face value, intended to provide an answer to the online harms of anti-democratic tendency discussed above, creating thereby a potentially dramatic shift in the relationship between the tech companies and criminal law. The existence of various expression-based offences now determines the nature of that relationship since they correspond with and delineate the online content now termed 'illegal' under the OSA. A clear incursion into online freedom of expression has therefore now occurred, *but* partly in the name of protecting democratic processes. The discussion below therefore questions the ability of the legal content duties to navigate a path between answering effectively to both interests.

But it is worth pointing out that the duties to protect users from harmful content are, as Coe has previously found, 'hard-edged' since, theoretically, they *must* be met,[135] whereas in contrast, the free speech duties imposed on some services are 'softer edged'.[136] So, very significantly, are those in relation to promoting 'content of democratic importance', since the OSA requires services only to 'take account of' or 'have regard to' them.[137] A mismatch arises therefore between duties to protect users from harmful content and duties to protect free speech, especially where it is viewed as of democratic importance. The scheme, taken at face value, is weighted against the latter concern. However, the discussion below indicates that weaknesses inherent in the operation of the duties to provide protection from online harms, combined with subjection of the duty of care to a non-transparent proportionality test,[138] clearly blunt the edge of such duties.

---

134 *White Paper* (n 3 above) para 6.15.

135 See eg OSA, s 10(4): 'The duties set out in subsections (2) and (3) apply across all areas of a service, including the way it is designed, operated and used as well as content present on the service, and (among other things) require the provider of a service to take or use measures in the following areas, if it is proportionate to do so—(a) regulatory compliance and risk management arrangements, (b) design of functionalities, algorithms and other features ...'.

136 Coe (n 41 above) 228.

137 OSA, ss 17(2), 19, 22, 7(5).

138 See eg OSA, s 10(4) (n 135 above).

Divisions between 'priority' and 'non-priority' illegal content: misalignment with curbing anti-democratic harms

What is 'illegal content'?

At the heart of the OSA framework lies reliance on the notion of translating criminal norms into regulation. Its regulatory response to the two categories of illegal content[139] – priority and non-priority – relies on findings by the tech companies in relation to regulated services that online content can be deemed 'illegal'. It is, therefore, necessary to consider the OSA's definition of illegality. 'Illegal content' is defined in section 59(2) as content amounting to a 'relevant offence'.[140] Section 192 provides that services must find illegality if they have 'reasonable grounds to infer' that the elements of the offence are made out,[141] and they do not have reasonable grounds to infer that a defence to the offence may be successfully relied upon.[142] Crucially, this turns on evaluations as to the state of mind of the person responsible for the content in terms of the *mens rea* of the relevant offence and as to whether they would have an available defence. One of the obvious problems in this position is that, in some instances, unless the service has information from which it can reasonably infer that the requisite state of mind was present or that a defence may be successfully relied on (which in itself may be of uncertain scope, such as relying on a 'reasonable excuse'),[143] it may not be possible to establish on the 'reasonable grounds' test that an offence could have been committed, meaning that the online content has to be disregarded,[144] even if the *actus reus* of the offence appears to be present. This places a requirement on the regulated services to anticipate illegality – and consequently remove content – only based on information reasonably available to them, meaning that what could be valuable extrinsic contextual information will inevitably be omitted from their assessments.[145]

139 'Illegal content' does not include relevant offences arising under the common law: OSA, s 59(4)–(5).

140 Ibid s 59(4), (5): content that is linked to priority or non-designated offences.

141 Including the *actus reus* and *mens rea* elements: ibid s 192(5), (6)(a).

142 Ibid s 192(6)(b).

143 See eg OSA, s 179(1)(d): 'the person has no reasonable excuse for sending the message'; this is not technically a defence since it is one of the elements the prosecution would have to prove, but it is a matter that the service in question would have to evaluate for the purposes of the OSA regulatory framework.

144 OSA, s 192(2) and (6)(b).

145 E Harbinja and N Ni Loideain, Policy Report: Making Digital Streets Safe? Progress on the Online Safety Bill (IALS and Aston University June 2023) s 3.3. See also E Judson et al, 'The bypass strategy: platforms, the Online Safety Act and future of online speech' (2024) 16(2) Journal of Media Law 336–357, 344–346.

This position clearly infuses uncertainty and variability into the responses of the in-scope companies. The sanctions regulated services are faced with, considered in the third and final section below, could mean that some services are likely to programme their algorithms to manage this risk, thereby erring on the side of caution when it comes to the retention and removal of certain (arguably) illegal content. This approach could lead to the filtering and removal of *legal* (but potentially harmful) online content if uncertainty as to the scope of an offence, such as section 127 Communications Act 2003,[146] or as to its application to particular content is evinced by some online services.[147] Censorship of content by those services might be the result. But, conversely, uncertainty as to whether content is within the scope of a particular offence, or whether a defence would apply, is likely to fuel arguments from several of the less compliant companies to the effect that failures to remove content were justified.

These weaknesses in the scheme for curbing the presence of illegal content on online platforms accord further support to the argument that a mismatch arises between the aspiration of the OSA to curb anti-democratic harms and the reality of the scheme it creates to do so. Reliance on the concept of 'illegal content', and on the ability of the tech companies to identify it effectively, clearly fuels the doubts expressed here as to the OSA's ability to navigate a path successfully between disproportionately curbing politically valuable free expression and addressing anti-democratic online harms. Given that many of the services are not minded fully or at all to comply with the OSA,[148] the latter concern is unlikely to prevail, bearing in mind that, if content is deemed legal, it would be subject to self-regulation only (if aimed at adults).

'Priority' and 'non-priority' illegal content: distinctions and differing duties

The distinction created by the OSA regulatory scheme between 'priority' and 'non-priority' illegal content creates, it is argued, a further flaw. There is a significant division in the scheme in terms of the demands of the duties relating to 'priority illegal content' (PIC) and non-priority 'illegal content' (IC). The argument below seeks to demonstrate that the designation of some online material as merely 'illegal' content leads to weaknesses in the ability of the OSA to place curbs on it. Thus, mismatches arise between the harm such content causes and the OSA response; as explored below, several of the anti-democratic harms

146 See the discussion in the subsection 'Online threats and abuse targeting politicians' below for criticisms of this offence.

147 Ibid.

148 See n 115 above.

identified in the first section of this article are designated *only* as non-priority content.

The key differences between the responses to PIC and IC are apparent from the illegal content duties under section 10,[149] applying to user-to-user services and, under section 27, to search services.[150] There is a general provision in section 10(2)(a) to the effect that user-to-user services must *prevent* users from encountering PIC. That wording indicates that the duties of the service providers are more onerous in relation to PIC, since it should be prevented from appearing on the service,[151] which appears to mean that it should be entirely excluded by design, or, if it appears on the service, under section 10(3), the service must minimise the length of time for which PIC is present and minimise its dissemination. But under section 10 the duties in relation to IC, in contrast to those applying to PIC, demand only that the services must 'swiftly'[152] remove IC, which seems to allow the services more flexibility in terms of the time that the content can remain available. It further appears that the content need not be excluded by the *design* of the service, since it is not required that users are *prevented* from encountering it.

As regards both user-to-user and search services the trigger for the removal of content also differs, depending on whether PIC or IC is in question, and again is more demanding in relation to PIC. The service must act to exclude content falling within the higher category, even where it has not been alerted to the presence of such content on the service. But, in contrast, it need respond to illegal content not within the 'priority' category only after the service has become aware of its presence by being alerted to it.[153] Sections 10 and 27 indicate that it

149 User-to-user services must comply under s 6(1) with the duties under s 10.

150 OSA, s 27(3) provides: 'A duty to operate a service using proportionate systems and processes designed to minimise the risk of individuals encountering search content of the following kinds—(a) priority illegal content; (b) other illegal content that the provider knows about (having been alerted to it by another person or become aware of it in any other way)'. The relevant codes have now been issued: see Ofcom, 'Illegal content codes of practice for search services'; Ofcom, 'Illegal content codes of practice for user-to-user services' – both issued 24 February 2025, in force 17 March 2025.

151 OSA, s 10(2).

152 OSA, s 10(3)(b).

153 Ibid s 10(3) creates: 'A duty [in relation to user-to-user services] to operate a service using proportionate systems and processes designed to—(a) minimise the length of time for which any priority illegal content is present; (b) where the provider is alerted by a person to the presence of any illegal content, or becomes aware of it in any other way, swiftly take down such content'. The equivalent provisions for search services in relation to illegal (not *priority* illegal) content arises in s 27(3)(b): 'other illegal content that the provider knows about (having been alerted to it by another person or become aware of it in any other way)'.

does not have to proactively seek out and remove such content as it would have to do with content in the higher category. The distinction created between PIC and IC is therefore crucial, since the latter is not subject to the highest OSA standard of protective intervention.

However, the notice and takedown provisions demand reliance on 'proportionate' processes only, meaning that PIC might remain available for significant periods if the provider was able to argue that seeking to identify it would involve a disproportionate outlay of resources, probably of especial advantage to smaller but in-scope companies. But the emphasis on proportionate responses in sections 10 and 27 might be even more problematic in relation to IC; the inability of a smaller service to employ sufficient moderators to ensure that such content is 'swiftly' removed under section 10,[154] or to minimise the risk of individuals encountering illegal search content, could be considered by Ofcom in determining that the response could be viewed as proportionate, even if the content remained available on some services for longer periods of time than on others. The service and Ofcom might take the view, in accordance with the priorities revealed by the categorisations in question, that resources should largely be devoted to addressing PIC.

These differences between the approaches to IC and PIC are of clear significance in terms of the scheme's ability to address the anti-democratic harms discussed in the first section. The two categorisations of illegal content lead to misalignment with the aim of curbing such harms. Most such harms are addressed by the OSA regulatory scheme as 'illegal' content only; therefore, they do not attract the highest level of protective intervention, indicating that curbing online expression of anti-democratic tendency was not in reality viewed as a high priority under the OSA, despite some appearances to the contrary. The apparently pro-democratic aspirations of the OSA are misaligned, therefore, with the decisions made as to the categorisations of types of content, given that the applicable duties differ in terms of the level of their demands. Below, the implications of this and, more generally, of the attempt to translate criminal offences into regulation, are analysed in relation to the anti-democratic online harms considered in the first section of this article. The three key examples are taken below.

154 That is implicit in s 10(4), which provides that the service should 'take or use measures in the following areas, if it is proportionate to do so'. One of the areas, in s 10(4)(e), is 'content moderation, including taking down content'. That can also be said of s 27(3)(b).

(i) The OSA response to online false information

If the OSA is viewed as an attempt to impose new governance standards on the tech companies to seek to provide protection for the health of the democracy, it follows that preventing the promulgation of mis- or dis-information online should be a high priority. Section 179 OSA repeals section 127(2)(a) and (b) of the Communications Act 2003 and creates a new, revised offence dealing with the sending of false communications.[155] But the question arises whether the online harms considered in the previous section[156] arising from the promulgation of false information, due to its anti-democratic impacts, are effectively answered to by this new offence. The *mens rea* of the section 179 offence includes: (i) knowledge of falsity, in that the defendant knew, rather than believed, the message to be false,[157] and (ii) that the defendant intended the message to cause non-trivial psychological or physical harm.[158] This is problematic for at least three reasons. Firstly, knowledge of falsity restricts the offence's ambit to dis-information only, which means that mal-information[159] is not covered.[160] This liability gap – which was raised by consultees to the Law Commission's proposals for reform,[161] and acknowledged by the Commission[162] – limits the scope of the regime since it does not account for the granularity of online communications.[163] Nevertheless, the Commission's solution – to recommend the creation of a further 'harm-based' communications offence – was not included in the OSA.[164]

---

155 S 179 covers such sending of messages online or offline, but makes particular provision for the online context in s 179(2): 'For the purposes of this offence an individual is a "likely audience" of a message if, at the time the message is sent, it is reasonably foreseeable that the individual—(a) would encounter the message, or (b) in the online context, would encounter a subsequent message forwarding or sharing the content of the message.' The false communications provision in s 1 Malicious Communications Act is also repealed.

156 See n 97 and associated text.

157 This is the same as the *mens rea* required for the s 127(2) Communications Act 2003 offence. See Law Commission, *Modernising Communications Offences: A Final Report*, HC 547, Law Com No 399, 20 July 2021, para 3.19, 79.

158 These do not operate in isolation but rather must be taken together: Law Commission (n 157 above) para 3.55, 90. The harm intended to be caused must therefore amount to such harm and be without a reasonable excuse.

159 True information deliberately used in a misleading way. See Council of Europe (n 90 above).

160 See Law Commission (n 157 above) and associated text re *Modernising Communications Offences*, para 3.44, 86.

161 Ibid para 3.43, 3.44, 86.

162 Ibid para 3.47, 87.

163 Ibid. See the argument made by Demos at para 3.27, 81.

164 It was argued that its lower threshold posed a risk to free speech. Eg see Carla Lockhart, MP for Upper Bann, HC Deb 19 April 2022, vol 712, col 117.

Secondly, falsity itself is often far from easy to determine,[165] posing significant difficulties for the Crown Prosecution Service (CPS) in seeking to establish the requisite knowledge. The task of distinguishing between trustworthy and false information is clearly fraught with difficulty since falsity may often be a matter of dispute,[166] although arguably possibly not as readily in the political context with which this article is concerned as in the medical or scientific one. Moreover, removing content expressing imprecise and ambiguous notions would be incompatible with international standards for restrictions on freedom of expression. The second aspect of the *mens rea* could compound this challenge for prosecutors, since it is not yet clear what 'non-trivial psychological or physical harm' means; Law Commission consultees found that it is hard to define with any precision.[167]

Thirdly, the threshold of seriousness is unclear. The Commission stated that to avoid over-criminalisation, by setting a low level of culpability, knowledge of falsity must be coupled with the intention to cause harm.[168] Setting this threshold was seen as crucial to prevent disproportionate interferences with free speech[169] but, since the offence is linked to the intention to cause harm, but does not refer to actual harm, it could be interpreted and applied over-broadly. But, conversely, it may prove hard to utilise in practice: the offence's two-pronged *mens rea* is combined with the general difficulty in determining falsity and non-trivial harm; thus the opacity of the threshold is likely to raise difficulties of proof, particularly in respect of borderline cases.[170] Due to such uncertainty, the capacity of this offence to combat the serious threat to democracies posed by the spread of false information online is called into question, as is its ability to strike an effective balance between addressing that threat and protecting free expression. The problems of curbing dis-information online via the OSA regulatory scheme, founded on the nature of this offence, are clearly compounded by its uncertainties and limitations.

---

165 This point was made eg by English PEN in Law Commission (n 157 above) para 3.25, 81.

166 See further on this issue, A Jungherr, 'Foundational questions for the regulation of digital disinformation' (2024) 16(1) Journal of Media Law 8–17.

167 Ibid para 3.45, 86–87.

168 It stated that so providing set a higher threshold than 'causing annoyance, inconvenience or needless anxiety' under s 127(2) Communications Act 2003: Law Commission (n 157 above) para 3.54, 90. See also *Harmful Online Communications: The Criminal Offences* (Law Commission Consultation Paper No 248 2020) para 6.45.

169 Law Commission (n 157 above).

170 Ibid para 3.53, 89–90.

The section 179 offence not only falls only within the illegal content category, but it is also, as discussed, uncertain of meaning and scope. That creates severe problems of interpretation[171] for the police, CPS and judges. But when false information arises in the form of illegal content, such problems also afflict the services and Ofcom in terms of identifying and 'swiftly' removing content arguably covered by section 179. Those problems are greatly intensified when it is borne in mind that once the section 179 offence is viewed as determining the ambit of a specific form of illegal content, it must – if the regulation is taken at face value – be responded to without all the additional aids that would be available to the prosecution when building a case, such as witness statements and evidence of the defendant's bad faith from other sources or media. Without the help of those aids, the services must in theory be satisfied, by considering the content alone, that the person posting it must have known that the message was false; they must also be satisfied that it is in fact false. They must further make an estimate as to whether the poster in question intended to cause non-trivial physical or psychological harm to a likely audience by posting the information and must make that determination by considering the information alone, since normally they would not have any further information as to the poster's intention in posting it.

One of the elements of the offence that the prosecution must prove (it is not a defence) is that of showing that the defendant did not have 'a reasonable excuse' for sending the message.[172] A moderator working for a tech company would usually be unlikely to be in possession of the information required to determine whether any excuse, reasonable or otherwise, was present. Clearly, in relation to online material, if uncertainty arises as to any one of the elements of the *actus reus* or *mens rea*, or as to the matter of finding a reasonable excuse, the content in question would be likely to be deemed legal. Therefore, no duty to remove it would be triggered, even if evidence *was* available indicating its falsity. The identification of content as falling within the scope of section 179 appears to present the tech companies with almost insuperable problems; those companies minded to engage in highly minimal compliance with the OSA regulatory scheme are likely to find it relatively straightforward to satisfy Ofcom that their failure to

171 Eg removing content arguably amounting to mis- and dis-information relating, for example, to misogyny or climate change, would not be required in relation to adult users if the ambiguities in the wording of the offence and defence in s 179, as discussed, rendered it possible on the facts that the offence would not be made out.

172 OSA, s 179(1)(d).

identify and remove arguably false information on a service was due to uncertainty as to whether the section 179 offence would apply.

There is therefore not only a mismatch between the dangers to democracy of the promulgation of false information online and the lack of priority given to addressing that harm under the OSA,[173] but also grave difficulties arise, clearly exploitable by the services, in seeking to strike a balance between preserving free expression online and curbing false information, since the scales are so heavily weighted against the latter concern. Clearly, the problems entailed in translating criminal offences into illegal content, are not confined to reliance on section 179. But the complexities and uncertainties of that section render this offence especially resistant to regulation under the OSA scheme, making it likely that false information could frequently remain available online since it would often be subject to self-regulation only.

(ii) Online threats and abuse targeting politicians

Rising and prevalent online abuse, threats and intimidation of politicians, discussed in the first section, has been termed a 'significant threat to democracy'.[174] The OSA itself creates a number of new offences that could potentially address this issue.[175] Section 127(1) Communications Act 2003 (CA) has already been found to encapsulate instances of cyber-bullying since it covers grossly offensive or

173 Possibly in recognition of this weakness, in the face of the current and growing concern as to the anti-democratic impact of false information online, s 152 places an obligation on Ofcom to establish an Advisory Committee on Dis- and Mis-information. But the potential influence the Committee's recommendations might eventually have is clearly open to question.

174 In a report from the Jo Cox Civility Commission, *No Place in Politics: Tackling Abuse and Intimidation* (Jo Cox Foundation 2024): see also n 195 below.

175 The range of relevant existing offences was thus expanded to include new ones, and certain offences were broadened, as recommended in 2021 by the Law Commission; the Commission made recommendations as to reform of the communications offences, in the 2021 *Modernising Communications Offences* report (n 157 above). See also notes 180, 181, 183 and associated text below.

menacing messages,[176] and that remains the case.[177] In default, at the time, of the more targeted offences now included in the OSA, it has been deployed on a number of occasions in relation to online threats to female MPs.[178] But its breadth and imprecision render it unsuitable as a means of creating an effective navigation between curbing online harms and the preservation of free speech.[179] It now, however, overlaps with the much more precisely worded and more serious new offence arising under section 181 OSA covering threatening

176 The Communications Act 2003, s 127(1)(a) provides:' Improper use of public electronic communications network: A person is guilty of an offence if he sends by means of a public electronic communications network a message or other matter that is grossly offensive or of an indecent, obscene or menacing character; or causes any such message or matter to be so sent'. In the leading case, *Director of Public Prosecutions (Appellant) v Collins (Respondent)* [2006] UKHL 40 (on appeal from [2005] EWHC 1308 (Admin)), it was relied on in relation to offline racist expression – phone calls – found to be 'grossly offensive'. But s 127 has more recently been applied to online expression: see Law Society of Scotland, '"Nazi pug" accused refused leave for Supreme Court appeal' (Law Society of Scotland 23 January 2019); *Scottow v CPS* [2020] EWHC 3421 (Admin) concerned s 127(2)(c) (which was also not repealed under s 189(1) OSA). S 1 Malicious Communications Act 1988 also covers a message which is (a) '(i) indecent or grossly offensive, or (b) any article or electronic communication which is, in whole or part, of an indecent or grossly offensive nature'. The offence is committed if 'his purpose, or one of his purposes, in sending it is that it should, so far as falling within paragraph (a) or (b) above, cause distress or anxiety to the recipient or to any other person to whom he intends that it or its contents or nature should be communicated'. The other parts of s 1 which covered '(ii) a threat; or (iii) information which is false and known or believed to be false by the sender' were repealed by s 189(2)(a) and (b) OSA.

177 Since, surprisingly, it remains unrepealed (see n 179 below).

178 In the *Peter Nunn* case (unreported, 29 September 2014), the defendant retweeted abusive and menacing Twitter messages from several Twitter accounts to Labour MP Stella Creasy after she campaigned to put Jane Austen on the £10 note. The City of London Magistrates' Court heard that Nunn had retweeted 'menacing' posts threatening to rape her and branding her a witch. District Judge Elizabeth Roscoe found him guilty of sending indecent, obscene or menacing messages under s 127 and jailed him for 18 weeks (see S Creasy, 'Twitter intimidation not taken seriously enough by police' *The Guardian* (London 29 September 2014). David Begley made references to rape in Twitter messages to Plaid Cymru leader Leanne Wood as she appeared on a TV debate about the referendum. Ms Wood represented the Remain campaign in debates during the build-up to the referendum vote. Begley admitted sending a communication conveying an offensive message – offensive comments posted on Twitter. He was sentenced at Swansea Magistrates' Court to 12 weeks' imprisonment (see 'Internet troll jailed over tweets to Leanne Wood' (*BBC News* 15 July 2016).

179 The Law Commission's recommendation to repeal s 127(1) due to those failings was not accepted: see Law Commission (n 157 above).

communications[180] – those sent in order to convey a threat of death or serious harm[181] – designed to capture *inter alia* online threats to rape, kill and inflict physical violence.[182] Sections 187 and 188 OSA also cover sending or sharing/threatening to share intimate images.[183] These OSA offences, due to the more precise wording of the *actus reus* in each instance, and the express inclusion of requirements to prove intention or recklessness,[184] are designed to address online harms while preserving free expression much more effectively than section 127 of the CA is able to do. Whether in practice they will be able to demonstrate a greater capacity to address the harm discussed above of targeting politicians, especially female MPs,[185] via online threats, is questionable, due to the general problems of deploying criminalisation of online content to found successful prosecutions.

Consequently, due to the introduction of these new offences in the OSA, there are now at least 12 targeted, communication-linked offences,[186] applying to persons posting online,[187] but unsurprisingly

180 OSA, s 181 was previously s 1(a)(ii) of the Malicious Communications Act 1988 which is now repealed.

181 Ibid s 181(c) provides that the person sending the message intends that a person encountering it (therefore not necessarily the intended recipient) would fear that the threat would be carried out or is reckless as to whether they would so fear.

182 Two further new offences arise targeting harmful online expression: s 183 OSA covers sending or showing flashing images to people with epilepsy intending to cause them harm ('epilepsy trolling') as a new offence, while s 184 creates an offence of encouraging or assisting serious self-harm with intent to do so.

183 'Cyber-flashing' and 'revenge porn' also appear to be covered. The provisions amend the Sexual Offences Act 2003 to insert ss 66A and 66B. Note also s 240(1) OSA and the Online Safety Act 2023 (Commencement No 3) Regulations 2024 SI 2024/31, reg 2.

184 OSA, s 187, which covers the sending of intimate images, including deep fakes (s 187(5)), provides (s 87(1)) that intention or recklessness must be proved as to causing alarm, distress or humiliation due to the material sent, or as to obtaining sexual gratification from sending it.

185 See the subsection 'Targeting politicians: cyber-bullying, online abuse, deep fake pornography' above.

186 They also arise under s 1 Malicious Communications Act 1988 and s 4A Public Order Act 1986. S 4A, as amended, has been found to apply to abusive forms of online expression. See eg *S v DPP* [2008] EWHC 438 (Admin). But s 127(1) of the Communications Act 2003 has been relied on in preference to s 4A in more recent instances: see notes 176–178 above.

187 These offences run alongside the more general expression-based offences, applicable online and offline, not necessarily targeting individuals, in relation to curbing pornographic expression, expression-linked counter-terrorism and hate speech offences.

not to the services themselves.[188] This expanded array of such offences provides, at face value, an enhanced opportunity to address a number of the harms set out in the previous section, including posting or threatening to post deep-fake pornographic images of female MPs,[189] or sending them rape threats.[190] But the problems that have constantly plagued prosecutions under such provisions, including lack of police resources to investigate and anonymous postings, cannot be addressed merely by expanding the area of criminalisation. This expanded web of offences *also* delineates some of the illegal content that under the new regulatory scheme should be taken down, once the provider becomes aware of its presence, thereby broadening the reach of the new regulatory framework, with the intention of partially remedying such prosecutorial deficiencies. But once again doubts arise as to how far that intention is likely to be realised.

Some particularly serious online threats or forms of stalking will now count as illegal content as falling within the public order offences of harassment, stalking and fear or provocation of violence,[191] which do count as 'priority' illegal content under schedule 7 OSA. But the use of threats falling within section 127(1) Communications Act 2003, which has been much more likely to be relied on in relation to forms of online bullying or abuse targeting politicians,[192] does not. A significant concern is that online threats covered by the new, more serious section 181 OSA offence also fall into the less serious category. Indeed, strikingly, none of the new communications offences[193] in the OSA itself, discussed above, fall into the priority category. So posts sending deep-fake pornographic images to female MPs, apparently covered by section 188 OSA, do *not* fall into the priority category, although, somewhat arbitrarily, the narrower, cognate offence under section 33 of the Criminal Justice and Courts Act 2015 does.[194] Therefore, one

188 The offences discussed apply to third parties posting on social media sites, but not to the platforms themselves, even where the intermediary has had notice as to the offending material but has failed to remove it. The current CPS guidance concerning prosecutions for offences perpetrated via postings on social media makes no mention, unsurprisingly, of prosecuting the intermediary itself: CPS, 'Social media: guidelines on prosecuting offences involving communications sent via social media' (31 January 2024).

189 See Sexual Offences Act 2003, s 66B, as amended by OSA, s 188.

190 Pursuant to OSA, s 181. See the subsection 'Targeting politicians: cyber-bullying, online abuse, deep fake pornography' above.

191 Public Order Act 1986, ss 5, 4A and 4 as amended,

192 See n 178 above.

193 See n 175 above and associated text.

194 It covers disclosing, or threatening to disclose, private sexual photographs and films with intent to cause distress. It falls under OSA, sch 7, para 30, as 'priority' illegal content.

of the key aims underlying the introduction of the OSA in relation to abusive or threatening posts and cyber-bullying, of obvious democratic significance in relation to targeting female MPs[195] and politicians generally,[196] is realised only in a somewhat weakened form as far as categorisations of content are concerned.

As far as the offences themselves are concerned, several problems arise when they are translated into IC to be addressed by moderators within online services. One arises due to the overlap between the potentially applicable offences. For example, a moderator could determine that particular content falls more readily within the offence under section 127 Communications Act 2003, rather than the more serious one under section 181 OSA, probably partly because it would normally be harder to make an estimate as to the *mens rea* of the person responsible for the content under section 181, given its requirements, previously discussed, of demonstrating intention or recklessness. Section 127 itself does not include a *mens rea* requirement, but, following *Director of Public Prosecutions v Collins*,[197] it seems (in relation to messages found to be grossly offensive) to consist of an intention that the message will be insulting or reckless as to that possibility.[198] The *actus reus* of section 127 is also couched in broader and much less precise terms than that of section 181; section 127 covers 'grossly offensive or menacing' messages, while section 181 requires that the message 'conveys a threat of death or serious harm'. Therefore, there appears to be a likelihood that online moderators might be tempted to focus on the more imprecise, less serious, offence under section 127. But its terms mean that it is also one that creates greater areas of uncertainty, meaning that the argument, especially from less compliant companies, that the content in question could be deemed legal in borderline cases would be more likely to be successful. Weaknesses, therefore, in terms of both the OSA categorisation of the offence and the nature of some of

195 Eg the Jo Cox Civility Commission (n 174 above) has highlighted research that found that 90% of female Members of the Scottish Parliament had feared for their safety, and almost 43 per cent of Welsh and Senedd Members had received a death threat. The case of Labour's Jess Phillips is highlighted (now Parliamentary Under-Secretary of State for Safeguarding and Violence Against Women and former shadow Minister for Domestic Violence and Safeguarding); she said she had received 600 rape threats via social media in one evening: S Laville, 'Internet troll who sent Labour MP antisemitic messages is jailed' *The Guardian* (London 10 February 2017).

196 See the first section of this article above.

197 *Director of Public Prosecutions v Collins* [2006] UKHL 40.

198 Ibid: 'a culpable state of mind will ordinarily be found where a message is couched in terms showing an intention to insult those to whom the message relates or giving rise to the inference that a risk of doing so must have been recognised by the sender' (para 11).

the offences applicable to online threats directed at politicians, might render this aspect of the regulatory scheme largely unfit for purpose, especially in relation to such companies.

(iii) Online hate speech and terrorism-linked content

As discussed above in the first section of this article, forms of hate speech and terrorism-linked expression can readily be viewed as presenting threats to the healthy workings of a democracy. The OSA regulatory scheme has, *prima facie*, a greater chance of addressing such content, as compared to content falling within the communication offences discussed, since much of it is deemed to be priority illegal content under the OSA.[199]

Posters on, for example, X can, at least in theory, be subject to criminal sanctions if forms of content attacking certain protected groups are posted that fall within the purview of the current offences covering hate speech[200] or expression-based counter-terror offences. A very wide range of offences linked to preparing acts of terrorism,[201] encouraging or glorifying terrorism[202] or searching for relevant information[203] are already available, applicable to individuals employing online or offline expression in the respects designated. The attempts in the

199 Schedules 5 and 7 OSA. New relevant offences may be introduced in future falling outside the definitions of terrorism and of hate speech but akin to both and likely then to be designated as PIC. In the wake of Axel Rudakubana's guilty plea to the murder of three young girls in Southport in 2024, the UK Government vowed to 'change the laws to ensure that lone killers with "extreme individualised violence" are charged with terrorism'. From March 2025 the Illegal Harms Content Codes will require platforms to take 'proportionate measures' to protect users from illegal content. See further n 150 above and n 226 below, and see E Sinmaz, 'UK experts warn of dangers of violent content being readily available online' *The Guardian* (London 21 January 2025).

200 Public Order Act 1986, pt 3 as amended. This is demonstrated by the CPS authorising Northamptonshire Police, in the wake of the July and August 2024 riots, to charge Lucy Connolly – the wife of a Conservative Councillor – with publishing material intending to stir up racial hatred, contrary to s 19 of the Public Order Act 1986, for posts on X in which she called for hotels housing migrants to be set alight and for 'mass deportation now'. However, despite this, X said the post did not breach its community standards. See CPS, 'Woman charged with publishing material intending to stir up racial hatred' (10 August 2024); A Stavrou, 'Tory councillor's wife "did not break X rules" with criminal social media post' *The Independent* (London 3 September 2024).

201 Terrorism Act 2006, s 5 as amended.

202 Ibid, s 1 as amended. S 3 covers the application of ss 1 and 2 to internet activity. See, for discussion, E Bechtold and G Phillipson, 'Glorifying censorship' in A Stone and F Schauer (eds), *The Oxford Handbook of Freedom of Speech* (Oxford University Press 2021) ch 28.

203 Terrorism Act 2000, s 58 as amended. The amendment relates to internet searches and downloads.

design of all these offences to create a balance between preserving expression and answering to democratic values by curbing its harmful manifestations[204] are the subject of an extensive and largely critical literature, usually attacking the over-breadth of such offences on free speech grounds.[205] The failures of these offences in that respect clearly also influence the ability of the OSA regulatory scheme to maintain that balance when the offences are in effect transmuted into illegal content subject to regulation.

Expression and information-linked counter-terrorist offences are listed in schedule 5 OSA and are within the priority illegal category category; therefore, if a service carrying terrorism-linked content is within scope, and is classed as Category 1, protection of users from that content should, if taken at face value, be at the highest protective level that the OSA scheme contemplates. That is clearly a welcome development in pro-democratic terms, given the strong evidence that terrorism relies heavily on online services to survive and thrive in terms of obtaining wide publicity for terrorist acts, and via the radicalisation of individuals.[206] Some terrorist groups are well versed in methods of relying on online services to further terrorist aims while protecting the identities of individuals, rendering the task of prosecutors much harder. Problems of obtaining evidence without endangering its sources also arise. Therefore, the use of this regulatory scheme to remove the content in question, or ensure that it does not appear online, is clearly necessary. However, this scheme is not without flaws in terms of that enterprise. Terrorist groups may turn to more obscure online services that are not within scope, or to ones outside Category 1.[207] Nevertheless, if moderators have some success in translating the various offences applicable to terrorism-linked expression online into PIC, despite the difficulties involved in making judgements about *mens rea* elements or defences discussed above, this new framework may mean that some terrorist-related content becomes less prevalent on a number of the larger platforms.

Similar points may be made as regards online hate speech since most of the key offences under part 3 of the Public Order Act 1986, as

204 Inconsistencies in relation to the balancing act are readily apparent, such as the availability of specific free speech defences in relation to religious and homophobic, but not racist, hate speech: Public Order Act 1986, ss 29J and 29JA.

205 Bechtold and Phillipson (n 202 above); E Heinze, 'Viewpoint absolutism and hate speech' (2006) 69 Modern Law Review 543–582; S Bishop, 'Should we hate hate speech regulation? The argument from viewpoint discrimination' (2024) 74(4) Philosophical Quarterly 1059–1079; I Hare and J Weinstein, *Extreme Speech and Democracy* (Oxford University Press 2009).

206 See nn 71–75 above and associated text. See also UNODC, 'The use of the internet for terrorist purposes' (UN 2012).

207 See nn 125–127 above.

amended, are listed in schedule 7 OSA, and are therefore priority ones. However, certain key offences – possession of racially inflammatory material or possession of material inflammatory on grounds of religion or sexual orientation – are missing from the list in schedule 7.[208] Therefore, those offences do *not* fall into the priority illegal category and receive only the lower level of intervention. The offences could be committed by individuals who download – as opposed to posting – material carried on the services with the requisite *mens rea*, if falling within the sections in question. In terms of this regulatory framework, the omission of these offences from the priority illegal category means that material inflammatory on the three protected grounds covered (race, religion, sexuality) could remain present online, and available for downloading purposes, for longer than material falling into the incitement to hatred categories, creating a flaw in the ability of the scheme to address the politically marginalising impacts of hate speech on certain groups.

A further flaw arises since, despite the evidence that much misogynistic material is available online, and is amplified by algorithms,[209] the continued lack of offences covering incitement to hatred on grounds of gender means that such material is completely unaffected by the new regulatory framework and remains purely a matter for self-regulation, at present. If specific women and female MPs are targeted, certain of the communication offences referred to above could apply, bringing the material within the category of illegal content. But online content that would be removed under the other incitement to hatred offences could remain available if the hatred incited did not relate to individuals, or to one of the three protected grounds, but only to women.[210] Online content of a general misogynistic tendency aimed at under-18s could be found to fall within the category of harmful content but not illegal content, which is discussed below, but that

208 Public Order Act 1986, ss 23 and 29G.

209 See S Weale, 'Social media algorithms "amplifying misogynistic content"' *The Guardian* (London 6 February 2024), referring to a report from researchers at University College London and University of Kent, called 'Safer scrolling: how algorithms popularise and gamify online hate and misogyny for young people'. A recent report from Vodafone found that there is 'a whole ecosystem of "outrage merchants" who are profiting from misogynistic content, with 52 per cent of boys engaging with [such] content': 'AI "aggro-rithms": young boys are served harmful content within 60 seconds of being online' (Press Release 2024).

210 Rishi Sunak has hailed the OSA as giving greater protection to girls online; he failed to mention this gap in the legislation: Weale (n 209 above). He reportedly said in relation to misogynistic material: 'I'm pleased we have passed the Online Safety Act over the last year and that means the regulator now has tough new powers to control what is exposed to children online.'

would be a matter of the interpretation of the relevant provisions (in particular, section 62).

In this context, it is of significance that the Hate Crime and Public Order (Scotland) Act 2021 was passed by the Scottish Parliament in 2021 and implemented in 2024; it allows for incitement to hatred on grounds of sex to be included at a later date.[211] As mentioned, section 59 OSA defines 'illegal content'. Under section 59(5)(c)(iv) illegal content, but not PIC, can be based on an offence arising in 'devolved subordinate legislation made by a devolved authority with the consent of the Secretary of State or other Minister of the Crown'. Thus, in future the anomalous situation could arise in which the tech companies should remove online content inciting hatred against women and girls as illegal content 'swiftly', even where, if the person posting the content was not in Scotland, they could not be prosecuted for such incitement. It may be found therefore that the anomalies and arbitrariness arising from choices made as to the content and protected grounds to be covered by hate speech offences are then compounded by choices made under the OSA regulatory scheme as to categorisations of content, discussed above.

The 'legal but harmful' provisions: finding a focus on combatting anti-democratic content?

Legal but harmful online expression could be caught by duties that are imposed on Category 1 regulated services to take down user-generated content that breaches the service's terms of service.[212] Additionally, such services must use 'systems and processes that allow users and affected persons' to report both 'relevant content' and persons they believe should be suspended or banned based upon the terms of service;[213] 'relevant content' is content that the services' terms of service designate as requiring action.[214] Terms of service are usually opaque and complex and may be changed by the service at any point, non-transparently, and without, it appears, external monitoring.[215] Leaving it up to the services to determine the content that is covered in their service terms, and relying on them to apply those terms

211 It was implemented on 1 April 2024. S 12(1) of the 2021 Act provides: 'The Scottish Ministers may by regulations add the characteristic of sex to the list of characteristics in [three] ... following provisions.' 'Sex' appears to denote 'gender'.

212 OSA, s 71(1), 72(3)(a).

213 Ibid s 72(5).

214 Ibid s 74(5).

215 Ibid s 72(5). Furthermore, it seems that users will have a role to play in identifying and notifying services of content that breaches the terms of service, albeit the extent to which this might occur is yet to become clear.

consistently, makes those services *de facto* arbiters of free speech, even in relation to adults.[216] Among the creation of various flaws, that means, in terms of the concerns of this article, that some speech of political value might be removed, while, conversely, some harmful expression of anti-democratic tendency, as discussed above, might not be, partly due to its designation as non-priority illegal content, and partly to uncertainty as to the scope and application of the relevant criminal offences when translated into IC.

Also, services within scope likely to be accessed by under-18s are required to assess the specific risks created by legal but harmful content through risk assessments; they are then required to mitigate the risks.[217] The relevant provisions may, however, give an impression of protecting under-18s that is not fully borne out by the reality, since certain services highly relevant to their experiences are not within scope; they include smaller platforms that may have a high proportion of under-18 users,[218] but also private communications, including email, SMS, MMS, private texting and one-to-one live aural communication services are exempt from the Act.[219]

The online harms identified in the first section of this article as anti-democratic are *not* prioritised under the legal but harmful provisions aimed at under-18s. The designations of harmful content as either 'primary priority content' or 'priority content' mirror the flaws already discussed as inherent in the provisions delineating the difference between 'priority illegal content' and 'illegal content'. In a similar fashion, the 'primary priority content'/'priority content' divide in effect creates de-prioritisation of content in the latter category, given that the duties in relation to content in the former one are more

216 See Coe (n 41 above) 235.

217 See OSA, s 9 (user-to-user illegal content duties which also cover under-18s), s 11 (user-to-user risk assessment duties which cover under-18s), s 12 (user-to-user safety duties in relation to under-18s). Duties as to children's risk assessments are set out in s 28; duties to protect under-18s' online safety are set out in s 29(2)–(8). Pt 3, ch 4, imposes duties on providers of regulated user-to-user services and regulated search services to assess whether a service is likely to be accessed by children. See also Ofcom, 'Protection of Children Code of Practice for User-to-user Services' (July 2025); Ofcom, 'Protection of Children Code of Practice for Search Services' (July 2025); Ofcom, 'Children's Access Assessments Guidance' (January 2025); Ofcom, 'Children's Risk Assessment Guidance and Children's Risk Profiles' (24 April 2025).

218 See nn 125–127 above and associated text; S Livingstone, 'Child online safety – next steps for regulation, policy and practice' (*LSE Media Blog* 22 January 2025).

219 OSA, sch 1.

likely to prevent under-18s encountering such content.[220] Once again, there is a lack of alignment between the designations of content and an objective of curbing content of anti-democratic tendency. Thus, forms of hate speech targeting under-18s but arguably falling outside the relevant offences, are within the latter category,[221] indicating that the silencing or marginalisation of some groups politically was not given a high priority under the OSA.[222]

False information targeting under-18s[223] and falling, or appearing to fall, just outside the scope of the section 179 offence, would not count as IC. But, surprisingly, such material does not fall within the two categories of 'primary priority content' or 'priority content' either. It could possibly be covered as 'non-designated content that is harmful to children'[224] and therefore would be covered by the less demanding age-group dependent duties.[225] However, the likelihood that it would be removed quite swiftly from a service would be low, given that it is not within the 'priority' category and could be disregarded if targeting older teens; moreover, it is unclear whether it is covered at all. This forms a clear gap in the OSA scheme; in relation to under-18s, therefore, harmful false information appearing to fall outside section 179 is able to flourish online, given that in most instances it would be unlikely to be caught by self-regulation.

---

220 OSA, ss 61 and 62. Under s 12(3) user-to-user services have: 'A duty to operate a service using proportionate systems and processes designed to—(a) prevent children of any age from encountering, by means of the service, primary priority content that is harmful to children.' Under s 29(3) search services have a 'duty to operate a service using proportionate systems and processes designed to—(a) minimise the risk of children of any age encountering search content that is primary priority content that is harmful to children'. But under s 12(3) and s 29(3) the duties in relation to other content, including priority content, deemed harmful to children, are dependent on the age group in question; older teenagers would probably not be covered by the provisions. Also, under s 12(3) user-to-user services must 'prevent' children accessing primary priority content, whereas in relation to other harmful content they have to protect children from encountering it, which could include using forms of age bars that might not provide full protection.

221 OSA, s 62.

222 Online hate speech directed at under-18s within certain minorities is a particular concern since clearly their age renders them more susceptible to its impact, which has also been found to spread to whole communities. See G Fulantelli et al, 'Cyberbullying and cyberhate as two interlinked instances of cyber-aggression in adolescence: a systematic review' (2022) 13 Frontiers in Psychology 909299.

223 This is a highly significant, under-researched issue. See EU Reporter Correspondent, 'Young are the "main targets of proponents of disinformation"' (16 July 2022).

224 OSA, s 60(4) which refers to s 60(2)(c).

225 See OSA, s 12(3) or s 29(3).

ROLE OF OFCOM, SANCTIONS, POLITICAL INFLUENCE: REALITY OF REALISING AN OBJECTIVE OF CURBING ANTI-DEMOCRATIC ONLINE HARMS?

The policing role of OFCOM

Weaknesses identified in the nature of the categorisations under the regulatory scheme, and the arbitrariness affecting the translation of criminal offences into IC are compounded in relation to the sanctions available under the OSA and the practicalities of enforcing the scheme. Ofcom, installed by the legislation as the online safety regulator, has a roadmap to regulation which is seeing it implement various aspects of the regime between the present time and 2026, including, *inter alia*, the introduction of codes of practice.[226] It has been provided with a broad suite of powers which could, if taken at face value, enforce the duty of care.[227] But deployment of these powers is likely to be highly time-consuming and resource-intensive since they involve a gradual escalation in terms of severity, which accords the provider a number of opportunities to evade compliance. Under section 130(1) Ofcom may give a 'provisional notice of contravention' relating to a regulated service to the service provider if there are reasonable grounds for believing that the provider has failed, or is failing, to comply with any enforceable requirement[228] that applies in relation to the service. That can be followed by a 'confirmation decision' if, *inter alia*, the service has failed to comply,[229] which specifies the steps to be taken within a certain time-frame.[230] Ofcom can issue a penalty notice and, if not complied with, a publication notice.[231] Ultimately, the service can theoretically be forced to comply with that notice via a civil action for an injunction.[232]

Of the sanctions available to Ofcom, one of the most draconian is that it can fine regulated services up to £18 million, or 10 per cent of annual

226 Ofcom has published its 'Illegal content codes of practice for search services' and 'Illegal content codes of practice for user-to-user services' (n 150 above). It has also published its Protection of Children Codes (see n 217 above). See also Ofcom, 'Important dates for online safety compliance' (17 October 2024).

227 OSA, part 7, para 19, ch 6.

228 Ibid s 131.

229 Ibid s 132(2).

230 Ibid s 133(4).

231 Ibid pt 7, ch 6: in particular s 150.

232 Ibid s 150(11): 'The duty under subsection (10) is enforceable in civil proceedings by OFCOM—(a) for an injunction, (b) for specific performance of a statutory duty under section 45 of the Court of Session Act 1988, or (c) for any other appropriate remedy or relief. Breach of the injunction could mean that senior executives of the relevant companies were found personally liable in civil courts.'

global turnover, whichever is higher, if they fail in their duty of care.[233] Furthermore, sections 144–148 provide for 'business disruption measures' that allow Ofcom to apply for a variety of 'restriction orders' if the regulated service has failed to meet certain conditions relevant to the restriction sought. They include the controversial power to require third-party companies to cease facilitating access to non-compliant services.[234] Section 110 also creates criminal offences, pursuant to section 109, for named senior managers of in-scope services in respect of failures to provide the requisite information needed to determine compliance.[235] Completely blocking a service from access by UK users (geo-blocking) is not expressly included in the OSA sanctions, but while geo-blocking it is not explicitly mandated, it could potentially be deployed to mitigate risks. If deployed, these sanctions could present real deterrents to the companies; but, clearly, much will depend on Ofcom's willingness and ability, given its under-resourcing, to utilise them in practice.

The extent to which Ofcom will, or can, police the scheme effectively *before* resorting to sanctions is largely a matter of speculation at present, but various black letter aspects of the scheme discussed may themselves create challenges for the regulator in undertaking that task in practice. Difficulties of interpretation in finding that commission of a criminal offence appears to have occurred to trigger duties in relation to illegal content have already been discussed, but the provisions, for example, determining territorial reach, also appear to present Ofcom with various difficulties of interpretation. The OSA has extra-territorial scope under section 4, applying to online content if the provider has 'links with the UK'; such links will be present if: there are a significant number of UK users; such users form a target market for the service;[236] or the service can be accessed in the UK, and there are 'reasonable grounds to believe there is a material risk of significant harm to UK

233 Ibid sch 13, para 4. See further Ofcom, 'Statement on online safety fees and penalties' (26 June 2025), which says (at ch 3, in particular para 3.82, 30) that: 'On balance, we have decided to determine QWR [qualifying worldwide revenue] using the worldwide revenue approach because it will help ensure QWR can provide an effective deterrent through a higher maximum penalty cap linked to the relevant parts of regulated services. While there are some arguments that would favour a UK revenue approach, in our view they are not sufficiently strong to justify adopting a UK revenue approach for both fees and maximum penalty caps.'

234 OSA, s 146 (access restriction orders). That could include requesting third-party companies to stop providing services or facilitating access to the non-compliant platform, meaning that it would be erased from search results, app stores, or links on social media posts.

235 Ibid pt 7, ch 4. For detailed discussion on the OSA's enforcement regime, see Law (n 47 above) 289–294.

236 OSA, s 4(5).

individuals' presented by content on the service.[237] The terms 'target market' and 'significant harm' are not defined; therefore consideration of the three methods of establishing the territorial reach of the OSA indicates that on slightly different facts a number of ways of establishing such reach could be available. In relation to harmful content but not illegal content, applying only to under-18s, the relevant duties would apply to certain services with only a small number of UK users only if '*significant* harm' can be demonstrated. That term could exempt such services in some instances, even if the content available on the service potentially falls within the provisions in question, since they merely use the term 'harmful'.[238] If so, self-regulation alone would continue as far as some content is concerned, since it would be outside Ofcom's reach.

Further, the OSA exhorts the services at various points to design systems to protect under-18s from certain forms of disturbing, but legal, content,[239] including forms of false information falling outside, or appearing to fall outside, the section 179 offence, but the companies would tend to prefer to maintain the current designs since they profit from the use of algorithms that prioritise more disturbing or controversial content. Leeway is accorded to both the services and Ofcom to determine what is meant by 'harmful'[240] in this context in order to ascertain whether a company has failed in its duties in relation to design. How far Ofcom will be prepared to adopt a stricter interpretation of 'harmful' than has been adopted by particular services under their own guidance is unclear at present. But given the fact that certain duties, such as duties to remove PIC from platforms, the subject of reasonably clearly defined offences, are less open to interpretation, Ofcom may prioritise such duties as easier to police and *less* open to creating conflict between itself and certain companies.

In general, then, Ofcom has some discretion as to enforcing/responding to certain aspects of the OSA, including: determinations as to the meaning of 'harmful'; in relation to non-priority offences of imprecise ambit, as discussed, and as to softer-edged duties, including the duty to promote 'content of democratic importance'. If other, more clearly defined, duties come close to overwhelming its resources these more ill-defined matters may be de-prioritised. Clearly, and very significantly, Ofcom is relatively under-resourced and funded compared to the regulated services, presenting a barrier to its effective operation as the regulator, although a somewhat opaque provision for Ofcom, at its discretion, to charge fees to the services is made in the

237 Ibid s 4(6).

238 Ibid ss 60–62.

239 Ibid s 12 applying to user-to-user services: 'Safety duties protecting children'.

240 Ibid s 234(2) '"Harm" means physical or psychological harm'.

OSA.[241] It may be concluded that, in relation to the online harms of anti-democratic tendencies that this article has focused on, not only are the provisions themselves open to criticism, as discussed, due to de-prioritisation of curbs on some such harms, but the monitoring and enforcement mechanisms are also of uncertain efficacy.

Democratic deficiencies at the heart of the OSA: governmental influence over the regulation and regulator

Under the OSA the Secretary of State for Culture, Media and Sport gains influence over Ofcom, and over the regulation itself.[242] The scheme accords a range of powers to the Culture Secretary,[243] enabling them to exert considerable political influence over the shape of the regulatory framework created,[244] while, in contrast, parliamentary influence is minimised.[245] That provides opportunities for lobbying by the tech companies away from parliamentary scrutiny, of obvious concern in relation to the apparently pro-democratic aims of the scheme.[246] As a related issue, it may also be asked whether the provisions allowing for political influence are in keeping with Council of Europe declarations as to the need for media regulators to remain at a distance from political power.[247]

At face value the governance of the regulatory scheme is not vested in a creature of government; Ofcom is deemed to be an independent

241 Ofcom has a wide discretion in this respect, but the amount to be paid is not determined by the OSA itself. See 'Duty to pay fees', s 84: '(1) OFCOM may require a provider of a regulated service to pay a fee in respect of a charging year which is a fee-paying year. (2) Where OFCOM require a provider of a regulated service to pay a fee in respect of a charging year, the fee is to be equal to the amount produced by a computation—(a) made by reference to—(i) the provider's qualifying worldwide revenue for the qualifying period relating to that charging year, and (ii) any other factors that OFCOM consider appropriate, and (b) made in the manner that OFCOM consider appropriate.'

242 Eg G Barkle and G Leacock, 'The Online Safety Bill: does it go far enough?' *The Barrister Magazine* 18 May 2021.

243 The post was previously held by Nadine Dorries, and then by Lucy Frazer, but Lisa Nandy (now in the Labour Cabinet as Culture Secretary) has taken over the post since a Labour Government was elected on 4 July 2024.

244 See nn 120, 125 above, and n 248 below.

245 See n 252 below and associated text.

246 See n 5 above and associated text.

247 See Recommendation CM/Rec(2018)1 (7 March 2018) as to 'concerns arising from pressure exerted on the media by political and economic interests' (para 8). See also the previous recommendation applying to the broadcast sector: Recommendation CM/Rec(2000)23 '[regulation] should be defined so as to protect [regulators] against any interference, in particular by political forces' (para 3).

regulator. However, the OSA scheme demonstrates an uneasy division of power between the Culture Secretary and Ofcom, which potentially could tend to ensure that its regulation does not clash in certain respects with the priorities of the regulated services. The ability of the Secretary of State, who may be susceptible to lobbying by the services, to determine Ofcom's 'strategic priorities' and ensure that they are in line with governmental ones, under a combination of sections 92 and 172 OSA,[248] is also of note. Further, there are a range of respects in which the regulator itself is open to political influence; the current Chair, a Conservative peer, is a political appointment who will continue in post until 2026.[249]

The IC codes of practice set out what the services can do to mitigate the risks of harm, and guidelines on Ofcom's approach to enforcement will emerge,[250] so they are likely to become a key determinant of the duties of the regulated services since they add crucial details to the duties placed on service providers. Therefore, this significant power has been devolved to an unelected body which will create a range of new rules, since, as discussed, the OSA duties are multi-realisable and open to a wide range of interpretations. Further, the Culture Secretary can direct that Ofcom modifies a code of practice in various circumstances,[251] while in contrast Parliament is accorded sparse and

248 Ofcom must have regard under s 92 to the Secretary of State's statement of strategic priorities – governed by s 172.

249 The current Chair (since 2022 for a four-year term) is Michael Grade, a Conservative peer, appointed by Nadine Dorries, and described in 2022 by Labour's culture spokesman, Chris Elmore, as a 'Conservative peer who is completely out of touch with the British public': see J Waterson, 'Government picks Tory peer Michael Grade to chair Ofcom' *The Guardian* (London 24 March 2022). Nadine Dorries as Culture Secretary at the time made the final decision as to the appointment in consultation with Boris Johnson. As an example of the potential problems that could arise in future, it was reported in 2021 that Paul Dacre, former editor of the *Daily Mail*, who was deemed unappointable when first interviewed for the post, would be able to reapply; see eg J Waterson, 'Paul Dacre "should be banned from reapplying" as Ofcom chair, says Tory MP' *The Guardian* (London 15 September 2021). Although he then withdrew, the incident was illustrative of the politicisation of this appointment. (It may be noted that in contrast the Chief Executive of Ofcom is currently Dame Melanie Dawes, previously a civil servant, since 2020.)

250 See n 226 above.

251 See, for example, OSA s 44(1) which provides: 'The Secretary of State may direct OFCOM to modify a draft of a code of practice submitted under section 43(1) if the Secretary of State believes that modifications are required for the purpose of securing compliance with an international obligation of the United Kingdom.'

probably ineffective oversight over the codes.[252] Accordingly, it can be argued that in purporting to address online harms, the OSA is bringing online regulation by Ofcom under the influence of political power by stealth in a manner that partially discards the checks and balances traditionally intended to ensure media regulator independence in relation to content.

The OSA's apparent allegiance to pro-democratic aims can thus be questioned, not only in relation to the duties themselves, as discussed, but also in terms of minimising parliamentary oversight over the scheme in key respects, while enabling interventions from government. The ability of the OSA scheme to navigate a path between preserving politically valuable expression and curbing damaging anti-democratic expression partly depends on the detail of the Codes of Practice and on Ofcom's willingness and ability to use its powers to hold the service providers to the OSA standards. Governmental influence over both the codes and the appointment of the Chair of Ofcom is therefore clearly a matter of concern.

CONCLUSIONS

This article concludes that the OSA measures are not robust enough to combat anti-democratic online harms. The new regulation leaves some leeway for self-regulation to continue post-OSA implementation due to the gaps and flaws in the measures discussed. Some providers and some content will in effect fall outside the scope of the provisions, either overtly, as is the case in respect of some smaller providers, or due to the uncertainties and opportunities for evasion discussed created by the statutory drafting as to the translation of a range of criminal offences into regulatory duties. The misalignment highlighted between the apparently pro-democratic intentions originally underlying the OSA and the decisions made as to the categorisations of types of content, also plays a part in creating room for the tech companies to manoeuvre. The continuing importance of political influence, creating, therefore, the potential for lobbying by the companies and problems of enforcement, compounds the statutory weaknesses affecting the

252 See ibid s 44 covering the Secretary of State's powers of direction, powers to modify a Code of Practice on various grounds, including that of national security, on the basis of a belief, not a reasonable belief. Under s 44(12) and s 45 a draft of the modified code after such a direction must be laid before Parliament, but the Secretary of State 'may, with OFCOM's agreement, remove or obscure information in the statement (whether by redaction or otherwise) in order to prevent the disclosure of matters that the Secretary of State considers would be against the interests of national security, public safety or relations with the government of a country outside the United Kingdom' (s 44(13)).

scheme for content moderation on pro-democratic grounds. Platforms obviously favour the leeway granted by self-regulation as enabling pursuit of a business model that amplifies user engagement, and therefore content, and so are likely to exploit these weaknesses to the full, although company compliance with this scheme will inevitably show clear variations between providers.

As a result, it is concluded that the OSA scheme fails to navigate a path effectively between curbing the outlined anti-democratic online harms and promoting democratic health via protecting freedom of expression. The contrast between the limitations discussed surrounding the duty to promote content of democratic importance and the more expansive approach taken – at face value – to curbing priority illegal content is illustrative of the underlying concerns that have affected the drafting and development of the statutory provisions. This situation has largely arisen since there is little evidence of a serious, consistent attempt to balance anti- and pro-democratic concerns. That appears to have been due to the awareness of resource constraints affecting Ofcom, to behind the scenes lobbying by service providers of officials or members of the Conservative Government[253] which pushed through the OSA and, simply, to the sheer difficulty of curbing online anti-democratic harms via national legislation in a way that does not undermine the democratic benefits provided by online services.

It is clearly the case that tackling those harms with which this article is concerned via such legislation in a way that still preserves the pro-democratic benefits of platforms is an arduous task, given that the industry, dominated by a handful of figures, is generally opposed to that endeavour.[254] Possibly there is no current or proposed national model of online regulation that could *fully* succeed in a complex task of that magnitude; but, as indicated above, the failings of the OSA discussed here exacerbate the difficulties inherent in attempting it. Global initiatives that are currently nascent, including opposing the current monopolising of the online space, might eventually command a greater measure of success.[255] Nevertheless, some strengthening of the current version of this national top-down regulatory scheme could be attempted in order to curb weaknesses and gaps enabling the

253 See n 4 above.

254 See J Borger, 'Elon Musk's beef with Britain isn't (only) about politics. It's about tech regulation' *The Guardian* (London 25 January 2025). He argues that 'Experts suspect X owner's interest in UK is to put pressure on authorities working to codify a new online safety law.'

255 See R Fay, 'A model for global governance of platforms' in Tambini and Moore (n 13 above) ch 14, 256–260, 262–266, proposing a body like the Financial Stability Board, the Digital Stability Board, which could disseminate global best practice that could be implemented at the national level.

continuance of the harms discussed if to an extent they were left to self-regulation only. Several of the criticisms advanced above, including as to categorisations of content and as to the problems of translating material within scope of criminal offences into illegal content, could provide a future basis for so doing without tipping it too far in the direction of invading online free expression, if the political will to do so is there, combined with acceptance of greater resource allocation to this endeavour.