Journal of Media Law

ISSN: 1757-7632 (Print) 1757-7640 (Online) Journal homepage: www.tandfonline.com/journals/rjml20

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

Stephanie Law

To cite this article: Stephanie Law (2024) 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, Journal of Media Law, 16:2, 263-300, DOI: 10.1080/17577632.2025.2459441

To link to this article: https://doi.org/10.1080/17577632.2025.2459441

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 12 Feb 2025.

Submit your article to this journal

Article views: 3218

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=rjml20

JOURNAL OF MEDIA LAW
2024, VOL. 16, NO. 2, 263–300
https://doi.org/10.1080/17577632.2025.2459441

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

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

Stephanie Law

Southampton Law School, University of Southampton, Southampton, UK

ABSTRACT

The Online Safety Act (OSA)[1] and Digital Services Act (DSA)[2] constitute the UK and EU's attempts to regulate the systems and processes of online service providers, and in so doing provide a framework to make the internet safer, mitigating risks of harm arising from the accessibility and dissemination of illegal and harmful content, while protecting users' fundamental rights and freedoms. While the legislative endeavours have similar aims, they adopt markedly different approaches, in relation to the concept of harm engaged and as regards their regulatory frameworks and approaches to multi-level governance. This paper outlines the enforcement frameworks envisaged under each piece of legislation, examining the role attributed to online platforms as private actors, the investigatory and sanctioning powers of national and supranational regulatory bodies, the scope for judicial intervention and opportunities for coordination between the European Commission and OFCOM in areas of common interest.[3]

ARTICLE HISTORY Received 6 July 2024; Accepted 10 January 2025

KEYWORDS Digital environment; Online Safety Act; Digital Services Act; private and public regulation; content moderation

Introduction

It is trite to state that the world in which we live is defined by the significance of its digital society or digital environment. One can hardly imagine going about our daily lives without engaging with digital platforms, offered by

CONTACT Stephanie Law s.law@soton.ac.uk

[1]Online Safety Act 2023 (hereinafter OSA).

[2]Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC [2022] OJ L277/1 (Digital Services Act; hereinafter DSA).

[3]European Commission, 'Daily News, 13 May 2024'3 <https://ec.europa.eu/commission/presscorner/api/files/document/print/en/mex_24_2568/mex_24_2568_en.pdf> accessed 01 June 2024.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

service providers.[4] Our offline and online lives are inextricably intertwined, and we are increasingly reliant on online platforms to engage with and participate in society. We use Google or DuckDuckGo to access information, identify the quickest route to work using Waze or Google Maps, buy and sell goods via eBay and apply for credit via banking apps; we communicate with friends (or strangers) via Facebook Messenger and Instagram, and use Twitter/X to inform ourselves and contribute to debates on issues of societal and public importance. These platforms offer benefits to individuals, providing increased opportunities for communication, integration, socialisation and general convenience. However, these benefits come with a range of nuanced risks of harm, arising from content encountered online. We may suffer economic loss, for example, as a result of online scams or fraudulent adverts; we may be subjected to exploitative and manipulative data practices leading to a loss of privacy, to misleading advertising or aggressive targeting, or we may be manipulated by deceptive or dark patterns which nudge our behaviour, distorting our decision-making and potentially encouraging us to act against our best interests.[5] Content can generate individual harm (for example, an individual encountering misinformation on vaccinations may result in that individual contracting COVID); this same content may also generate societal harms (facilitating the development of an anti-vaccination eco-system which threatens global health). Content may generate harms that are physical (for example, messaging that promotes suicide), or psychological (for example, abusive content related to someone's race or religion); some content generating harm may be unlawful (for example, indecent images of children or terrorist material) while others may be lawful but nevertheless generate harm (for example, 'gore' content or content that encourages self-harm or body dysmorphia).[6]

The Online Safety Act (OSA)[7] and Digital Services Act (DSA)[8] constitute the UK and EU's attempts to regulate the systems and processes of the digital environment; in so doing, they aim to establish a framework

[4]In this paper, the concept of 'service providers' is used interchangeably with 'digital platforms' and 'online platforms'.

[5]These issues traditionally having given rise to violations of the protection of personal data under Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [2016] OJ L119/1 (General Data Protection Regulation; hereinafter GDPR), or of consumer law, including, for example, Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council [2005] OJ L149/22 (the Unfair Commercial Practices Directive).

[6]All of these examples can be found in OFCOM, 'How People are Harmed Online: Testing a Model from a User Perspective' (2022) <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/how-people-are-harmed-online/> accessed 22 October 2024.

[7]Online Safety Act 2023 (hereinafter OSA).

[8]Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC [2022] OJ L277/1 (Digital Services Act; hereinafter DSA).

to make the internet safer, mitigating risks of harm arising from the accessibility and dissemination of illegal and harmful content, while protecting users' fundamental rights and freedoms. While both pieces of legislation seek to satisfy similar aims, they adopt markedly different approaches, not only in relation to the concept of harm engaged in each but also as regards their regulatory and enforcement frameworks and approach to multi-level governance. This paper aims to outline the enforcement frameworks envisaged under each piece of legislation, looking firstly at the role attributed to online platforms as private actors under the DSA and OSA, and furthermore at the role of national and supranational regulatory bodies, including the National Digital Services Coordinators, the European Board of Digital Safety and the European Commission under the DSA, and OFCOM under the OSA. It will unpack the investigatory and sanctioning powers attributed to these entities, the scope for judicial intervention and the opportunities for coordination between the European Commission and OFCOM in areas of common interest, as envisaged in an administrative agreement signed in May 2024.[9]

In order to contextualise these enforcement frameworks, the paper begins by outlining the recent legislative endeavours adopted within the EU and UK in the form of the DSA and OSA, and their shared endeavours to the ends of making the digital environment a safer place. Thereafter, the nature of the private entities controlling the digital environment, ie service providers, and their role in moderating online content, whether legal or illegal, is set out. Herein, the challenges arising in this framework of private or self-regulation are identified. Against the EU and UK legislative background, the obligations established on service providers in the OSA and DSA are examined, and an attempt is made to unpack the relationship between public and private regulation. The paper then moves on to examine the procedural framework by which compliance with the OSA and DSA will be enforced. Here, the added value of comparing the enforcement regimes will be elaborated upon; such an analysis allows for a clearer framing of the obligations and responsibilities of online service providers in respect of illegal and harmful content and offers a deeper understanding of what they should be doing and how they should be doing it. Examining the roles to be played by national and European authorities in relation to enforcement of service providers' compliance with the relevant legislation allows for the regulatory approaches of the EU and UK legislators to be further unpacked and offers insights into broader policy approaches as regards the public/private divide in the realm of content moderation under both regimes.

[9]European Commission, 'Daily News, 13 May 2024' 3 <https://ec.europa.eu/commission/presscorner/api/files/document/print/en/mex_24_2568/mex_24_2568_en.pdf> accessed 01 June 2024.

Legislative developments to the ends of a 'safer' digital environment

Both the OSA[10] and DSA have been advanced as legislative developments that will make the digital environment a safer place, where 'what is illegal offline is illegal online'.[11] Nevertheless, both pieces of legislation have generated criticism on various fronts, as will be explored further below, and have been subject to considerable amendments since their introduction before the UK and European Parliaments. The DSA replaces the e-Commerce Directive,[12] and finds its origins in the need, identified in the 'Shaping Europe's Digital Future' paper,[13] for a review and modernisation of the rules governing online platforms. Taking the form of a regulation,[14] the DSA became directly applicable across the EU from 17 February 2024.[15] The OSA stems from the Online Harms White Paper[16], which recognised the need for greater regulation in the face of the 'growing evidence of the scale of harmful content and activity that people experience online';[17] the OSA essentially also replaces the e-Commerce Directive, in the legal jurisdictions of the UK, post-Brexit. Having been introduced to Parliament in May 2021, and undergoing a number of significant amendments, the OSA finally received Royal Assent in October 2023 and came into force on the same day; it is expected that the Online Safety Act will be supplemented by various pieces of secondary legislation, codes of conduct and guidelines to be introduced in the coming months. It is worth noting that the push within the EU and the UK to enhance internet governance and regulation are also reflected in legal developments within other domestic legal

[10]Department for Science, Innovation and Technology, 'Press Release: Britain makes internet safer, as Online Safety Bill finished and ready to become law' 19 September 2023 <https://www.gov.uk/government/news/britain-makes-internet-safer-as-online-safety-bill-finished-and-ready-to-become-law> accessed 26 May 2024.

[11]Department for Science, Innovation and Technology, 'Press Release: Britain makes internet safer, as Online Safety Bill finished and ready to become law' ibid, quote of Secretary of State, Michelle Donelan. A similar sentiment has been expressed by European Commission President, Ursula Von der Leyen; European Commission, 'Press Release: Digital Services Act: Commission welcomes political agreement on rules ensuring a safe and accountable online environment' 23 April 2022 <https://ec.europa.eu/commission/presscorner/detail/en/IP_22_2545> accessed 26 May 2024.

[12]Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L178/1.

[13]European Commission, 'Communication on Shaping Europe's Digital Future' February 2020.

[14]Consolidated Version of the Treaty on the Functioning of the European Union (TFEU) [2012] OJ C326/47, Art 288. Regulations, aiming to ensure the uniform application of Union law across the Member States by replacing incompatible national laws, are generally applicable, binding and of direct effect against private individuals, Union institutions and the Member States from the date on which they enter into force.

[15]DSA, Art 93.

[16]Secretary of State for Digital, Culture, Media and Sport, and Secretary of State for the Home Department, 'Online Harms White Paper' (CP 2019) 57.

[17]ibid 12.

orders, including France,[18] Germany[19] and the United States,[20] towards enhanced online safety.

In the EU context, it is also worth noting that two pieces of legislation sit at the heart of the EU's endeavours to establish a legislative framework for the regulation of the digital single market: the DSA[21], which is the focus of this paper, and the Digital Markets Act.[22] Together, and alongside other legislative endeavours,[23] the DSA and DMA aim to provide for a harmonised set of rules governing digital services; these rules aim to facilitate fair, safe and contestable online spaces, and a level playing field within the internal and global market, to the 'benefit of the Union's economy as a whole and ultimately of the Union's consumers'.[24] The Commission envisages that together these Regulations will provide for a 'high global benchmark'[25] by updating pre-digital economy legal rules to be fit for the online environment and establishing clear and harmonised obligations on platforms, thus facilitating legal certainty while reducing compliance costs by negating fragmentation arising as a result of piecemeal regulation at the national and EU levels. Underpinning the need for the rules found in the DSA and DMA is a recognition of the market (but equally, social and political) power of the largest

[18]Loi n° 2021-1109 du 24 août 2021 confortant le respect des principes de la République, Articles 36 to 48, amending existing legislation including the Criminal Code, Code of Criminal Procedure and Consumer Code.

[19]Gesetz zur Verbesserung der Rechtsdurchsetzung in Sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – 'NetzDG') (1 September 2017). See U Kohl, 'Platform Regulation of Hate Speech – A Transatlantic Speech Compromise?' (2022) 14 *Journal of Media Law* 1.

[20]Legislative proposals have been advanced at both the federal – see for example, the Kids Online Safety Act S1409, introduced in the 118th Congress of 2023–24 – and at the state level – see for example, California's Age-Appropriate Design Code Act (AB 2273); the enforcement of the latter, due to take effect on the 1 July 2024, was enjoined by a preliminary injunction issued in September 2023 (*NetChoice, LLC v. Bonta*, No. 22-cv-08861 (N.D. Cal.)), a decision that has since been appealed to the Ninth Circuit by the State of California).

[21]DSA, Art 93; the DSA, as a regulation, became directly applicable across the EU from 17 February 2024.

[22]Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 [2022] L265/1 (Digital Markets Act; hereinafter DMA). The DMA, as a regulation, became directly applicable across the EU from 2 May 2023.

[23]Including, amongst others, for example, the consumer law *acquis*, the Treaty provisions on competition law and the GDPR, as well as recent Commission proposals developing a body of EU data law. The latter would include European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)' COM/2022/68 final; European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act)' COM/2021/206 final; European Commission, 'Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on EU Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act)' COM/2020/767 final – see further, Michèle Finck, 'The Maturation of European Data Law: From Fundamental Rights to Economic Rights' in Jeremias Adams-Prassl *et al* (eds), *The Internal Market Ideal* (OUP 2024) 39.

[24]DMA, Recital 8 and Art 1(1). Considerable reference is made within the DMA to the regulation of digital services, and the interrelation of the two regulations; see also, DSA, Art 1(1).

[25]European Commission, 'Europe Fit for the Digital Age: New Online Rules for Platforms' <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act/europe-fit-digital-age-new-online-rules-platforms_en#small-online-platform-scaling-up-in-the-eu> accessed 2 June 2024.

online platforms, characterised as 'gatekeepers' under the DMA.[26] While the focus of this paper it not on the DMA, it can be said that service providers, including online platforms, have been empowered to play a gatekeeping role in relation to internet governance, related to but distinguishable from the notion of gatekeeper as established in the DMA.

The role of private service providers in promoting a 'safer' digital environment: a brief introduction to content moderation

As set out in the introduction, the online environment and development of digital technologies offer a breadth of opportunities for societal engagement and interaction that are almost unimaginable in an offline world.[27] While the emergence of the digital environment, in particular, digital platforms, has allowed for an exponential rise in the sharing of ideas and information and the offering of products and services, potentially accessible to all, such opportunities also generate risks of harm. For example, unsafe products may be sold to consumers in breach of consumer law, or fake designer products may be sold in breach of IP and copyright law via online e-commerce platforms; hate speech or disinformation may be uploaded to discussion forums and incite violence or influence democratic participation, while images that promote self-harm or suicide may be posted on social media platforms. Some of this content may be illegal (for example, terrorist material which may incite violence or drive radicalisation) while other content may be lawful but nevertheless generate risks of harm (including for example, images that promote disordered eating). The digital environment may well offer positive outcomes, as outlined above; however, they undoubtedly also lead

[26]Under the DMA, these actors may be designated as 'gatekeepers' per Art 1, where they have a 'significant impact on the internal market', provide core platform services and 'enjoy an entrenched and durable position' (see also DMA, Recital 6). The DSA itself does not utilise the term 'gatekeeper' however, it is recognised at DMA, Recital 59, that gatekeepers often provide core platform services, thus it is evident there is clear overlap between the two.

[27]One might attempt to draw an analogy, for example, between eBay, as an online, customer to customer marketplace, and the 'buy and sell' noticeboards that might once have been found in a local newsagents or supermarket; similarly, politicians and tech leaders, and even the US Supreme Court (see *Packingham v North Carolina*, 137 S Ct 1730, 1737 (2017)), have been quick to draw analogies between online platforms, including social media forums, as digital iterations of the physical public square. See, for example, Elon Musk's 27 March 2022 tweet, a month before he announced his intention to acquire the platform and take it private, in which he espoused 'Given that Twitter serves as the de factor public town square … '. On the 'misguided and misleading' public square analogy, see Mary Anne Franks, 'Beyond the Public Square: Imagining Digital Democracy' (2021) *The Yale Law Journal Forum* 427, 428–429. Franks' challenge to the analogy lies not only in the private, for-profit nature of social media platforms but in its basis in the notion of the 'unregulated' public town square; rather Frank argues that the public square has aways been 'selectively regulated', 'reinforc[ing] legal and social hierarchies of race, gender, class, and ability rather than foster[ing] radically democratic and inclusive dialogue'.

to a diversity of disputes, political, ethical, societal and indeed legal, arising between different actors.[28]

In order to deal with the risk that individuals may access or be confronted with illegal or harmful content, there has been a turn to service providers to take on the role of 'content moderators'. Content moderation has been defined as 'the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse'.[29] In this role, service providers become gatekeepers of online content through the measures taken, or not, in the content moderation process. They decide what content can be accessible online, who can share content, who can access it, and whether it should be possible to monetise its sharing. Fundamentally, online platforms determine whether content is accessible; that is to say, they exercise control over its visibility by promoting or demoting such content to users. With its focus on the privatisation of public adjudicative functions, this paper identifies and analyses two sets of challenges; there are numerous others.[30]

The first concerns the private character of digital platforms. Firstly, digital platforms are private actors, for-profit and privately owned; their business models – largely advertising-driven, engineered to derive personal data from individuals[31] and engineer engagement, often via the use of algorithms and recommender systems – serve private, not public or societal, interests. The very private nature of the service provider itself generates a multiplicity of further challenges. Content moderation by service providers entails that the decisions made by private actors may impact individual (and societal) rights and interests. As such, private actors are engaged in an otherwise (though not exclusively) State function: the balancing of rights and freedoms.[32] If

[28]Goanta and Ortolani distinguish '(i) conflicts between users and the platform: users disagree with platform actions (e.g. taking down their content) or policies (e.g. nudity policies); (ii) conflicts between different users: users disagree between themselves, whether bilaterally (e.g. making disparaging or misleading statements about a business competitor …) … or multilaterally (e.g. 'cancel culture' practices …) … (iii) conflicts involving third parties: content posted on social media platforms may infringe the rights of non-users … '; Catalina Goanta and Pietro Ortolani, 'Unpacking Content Moderation: The Rise of Social Media Platforms as Online Civil Courts' in Xandra Kramer *et al* (eds), *Frontiers in Civil Justice* (Elgar 2022) 192, 198–199. Third party actors might include, for example, consumer protection associations, or child protection charities, who might request that content is removed.

[29]James Grimmelmann, 'The Virtues of Moderation' (2015) 17 *Yale Journal of Law and Technology* 42, 47. Content moderation is defined in the DSA at Art 3(t), while no definition is provided in the OSA.

[30]Including, for example, allegations of psychological harm incurred by content moderators (often outsourced) in the processes of content moderation. Indeed, broad criticism has been levelled against digital platforms and service providers engaged in content moderation, stemming from politicians, tech leaders, academics, international organisations, and human rights bodies.

[31]On the challenges arising in relation to the unprecedented power attributed to 'big tech' companies, their promotion of surveillance capitalism and its use as a tool to shape, influence and control markets and societal behaviour in a way that threatens autonomy, privacy and democracy – see further, Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs 2019).

[32]The multilevel balancing of rights and freedoms in the EU context, save certain absolute rights, is inherent in the Charter of Fundamental Rights of the EU (re freedom of expression, see CFR, Art 52) and the European Convention on Human Rights (re freedom of expression, see ECHR, Art 10(2)).

Twitter/X decides to remove a post of one of its users, it may be seen to constitute an interference with that user's freedom of expression.[33] It can be seen, as Ortolani notes, as a reflection of the 'privatisation of public adjudicative functions'.[34] In this sense, it can be taken for granted that 'content moderators act in a capacity very similar to that of judges'[35]; that is to say, service providers, in engaging in active content moderation, apply their own rules[36] to individual cases over and over again to decide whether or not content should appear on their platform.[37] This may be done by virtue of the platform's *ex ante* moderation or on the basis of a complaint made by a user in respect of content they have encountered on the platform.[38] These decisions may be right or wrong and can be subject to appeal; in this sense, content moderation is deemed to achieve accountability by attributing procedural rights to its users.[39] The online platform determines not only whether the content remains on its platform[40] but also determines its accessibility, via the promotion or demotion of content.[41] On this approach, online platforms take decisions on an individual, case-by-case basis; as these decisions are made over and over again, the process of content moderation results in 'the aggregation of millions of daily paradigm cases', which Douek likens to a 'bureaucratic model';[42] Douek – who refers to this approach as providing the 'standard picture' of individual content moderation – argues that it is one which does not describe the whole picture.[43] It is, she argues, 'outdated and incomplete', particularly as it focuses on whether the platform makes the right or wrong decision (one which can be challenged if wrong) in individual cases on

[33]This role for private actors, and in particular, digital platforms is not necessarily novel; see for example, the European Court of Justice's 'right to be forgotten' case law, arising in respect of GDPR, Art 15 (including Case C-131/12 *Google Spain* EU:C:2014:317, Case C-136/17 *Google v CNIL* EU:2019:773 and Case C-507/17 *Google v CNIL* EU:C:2019:772).

[34]See Pietro Ortolani, 'The Digital Services Act, Content Moderation and Dispute Resolution' (2022) SSRN <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4356598> accessed 15 May 2024, 3.

[35]Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 *Harvard Law Review* 1598, 1647. Klonick examines the approach of Facebook from a US perspective and in light of the First Amendment, asserting that these 'private content platforms' must be understood as systems of governance, 'now responsible for shaping and allowing participation in our new digital and democratic culture, yet they have little direct accountability to their users', 1599, a competence that derives from s 230, Communications Decency Act, which attributes to service providers immunity from liability in respect of content, generated by users and posted on their platform.

[36]Marjorie Heins, 'The Brave New World of Social Media Censorship' (2014) 127 *Harvard Law Review* 325, 326; Heins describes the process of Facebook's appeal mechanism as 'mysterious' and notes 'we do not know the details' of their internal policy.

[37]Evelyn Douek, 'Content Moderation as Systems Thinking' (2022) 136 *Harvard Law Review* 526, 537–538.

[38]Klonick (n 35) 1636–1638.

[39]Douek (n 37) 531.

[40]A consideration that Douek notes derives from analogies drawn with First Amendment challenges but which also attributes significance to high-profile decisions; Douek, ibid 565.

[41]Douek, ibid 546–548.

[42]Douek describes this 'standard picture' as 'privatized hierarchical bureaucracy that applies legislative-style rules drafted by platform policymakers to individual cases and hears appeals from those decisions', likened to offline decision-making before national courts, at 535; Douek ibid 535 and 538.

[43]Douek ibid 535. For a response to and critique of Douek's paper, see K Klonick, 'Of Systems Thinking and Straw Men' (2023) 136 *Harvard Law Review Forum* 339.

whether content should be removed.[44] Instead, in Douek's alternative vision of the governance of online speech, she advances a 'systems-based' approach. This approach, she argues, can overcome some of the challenges or 'blind spots'[45] with the 'standard picture' approach. Firstly, it engages the breadth of actors and institutions engaged, and often hidden, in content moderation, acknowledges the increasingly important relationships between private actor online platforms and governments,[46] and the trade-offs to be made in the institutional design of the moderation regimes.[47] While this paper affords little scope to delve in depth into Douek's model of content moderation, it is worth highlighting its parameters. Content moderation must, she argues, be conceived as dynamic and complex, concerned not only with decision making in individual cases; instead, she asserts, it must encompass a 'proactive and continuous form of governance',[48] one which 'focuses on the *ex ante* institutional design choices involved in creating a system of mass administration, rather than *ex post* individual error correction'.[49] Content moderation's goal should not necessarily only be concerned with the identification of norms that govern online speech or the 'application of a fixed rule to a specific piece of content'[50] instead, it must be concerned with the design of institutional mechanisms that encompasses problematic cases that fall outside the standard paradigm case, which encourages discourse and dialogue on values underpinning the design, implementation and enforcement of moderation regimes, so as to facilitate the meaningful accountability of online platforms.[51]

As Douek, and others do, it is necessary to also acknowledge that on a practical level, public adjudicatory institutions, like courts, do not have the capacity to deal with such a deluge of decision making, particularly as much of the content moderation undertaken by service providers is automated[52] (though certainly Meta does engage, often third party, human content moderators).[53] Facebook, for example, removes hundreds of thousands of posts a day, and thus certainly 'makes' more such decisions on a daily basis; of almost 70,000,000 pieces of content removed, over

[44]Douek (n 37) 538.
[45]ibid 539.
[46]For example, in countering child sexual abuse material.
[47]Douek (n 37) 539ff; such trade offs include, for example, increased scope for 'wrong' decisions to ensure decisions are made in a timely manner, the parameters of what is timely being different in an online than offline context, at 549.
[48]Douek ibid 531.
[49]ibid 533.
[50]ibid 607.
[51]ibid 533.
[52]Generating – from a technical perspective – the 'black box' problem and challenges of transparency and accountability; Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* (HUP 2015).
[53]See Douek (n37) 544.

67,500,000 were removed by automated means.[54] Certainly national courts have neither the capacity nor resource to deal with such a volume of cases and indeed, it might be asserted that the capacity of service providers to offer such decision making rather facilitates access to justice.[55] That is to say, when a user encounters content on a platform, they can make a complaint to the platform itself; that platform will then decide on whether that content should remain accessible. If that decision is wrong, the poster of the content can challenge that decision. This content moderation process allows for the resolution of what are necessarily disputes, albeit of a low monetary value, which arise with such frequent occurrence that they could never be resolved before national courts.

To the extent that content moderation entails such functions, one must question whether digital platforms are held to certain standards of 'justice' or whether they enjoy a freedom from the requirement of such compliance.[56] These questions might arise in relation to procedure and substance. It is debatable to what extent service providers, as private actors, are required to ensure due process and access to appeal mechanisms, for example, or whether they are bound by human rights standards in the same way that these rules bind States;[57] inherent here is the freedom of expression question.[58] The DSA recognises that content moderation processes of online platforms may generate conflicts with fundamental freedoms,[59] though it does not provide for procedural rules through which users can make complaints before national courts or administrative authorities (this being for the Member States themselves in line with the principle of national procedural

[54]Meta, 'Regulation (EU) 2022/2065 Digital Services Act: Transparency Report for Facebook (Period 01 October 2023 to 13 June 2024)' <https://transparency.meta.com/sr/dsa-transparency-report-apr2024-facebook> accessed 02 July 2024.

[55]Goanta and Ortolani (n 28) 210.

[56]Whatever we may take this to mean; for the purposes of this paper, it is impossible to unpack 'State-based' standards of justice in depth, or even whether we should transplant such concepts to a 'stateless' context; See Pietro Ortolani, 'The Three Challenges of Stateless Justice' (2016) 7 *Journal of International Dispute Settlement* 596, 597–601.

[57]For the argument that private actors, including online platforms, should ensure compliance with international human rights law when engaged in an adjudicatory role, including content moderation, see Report of the Special Rapporteur to the Human Rights Council on online content regulation, A/HRC/38/35 (2018).

[58]The freedom of expression and free speech discussion necessarily must fall outside of the scope of this paper; see however Barrie Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation' (2020) 43 *Fordham International Law Journal* 939 and Uta Kohl, 'Towards a Regulatory Theory of Platform Rule: Corporate 'Sovereignty' through Immunities' (2025) 56 *St Mary's Law Journal*.

[59]DSA, Recital 3: 'Responsible and diligent behaviour by providers of intermediary services is essential for a safe, predictable and trustworthy online environment and for allowing Union citizens and other persons to exercise their fundamental rights guaranteed in the Charter of Fundamental Rights of the European Union (the 'Charter'), in particular the freedom of expression and of information, the freedom to conduct a business, the right to non-discrimination and the attainment of a high level of consumer protection' and Recital 22: 'The removal or disabling of access should be undertaken in the observance of the fundamental rights of the recipients of the service, including the right to freedom of expression and of information' and DSA, Art 34(1).

autonomy); the DSA does establish procedural rules where a user triggers the Article 16 DSA notification procedure, and for example, requirements as regards the public accessibility of platforms' terms and conditions.[60] These procedural rules and the way in which they operate to bind online platforms are examined further in later sections of the paper.

Not only is content moderation done by private actors, but it may also be done – particularly concerning potentially harmful but not necessarily unlawful content – on the basis of privately-made law.[61] That is to say, while certain content moderation decisions are made on the basis of legally-binding rules, including for example, criminal law,[62] IP and copyright law,[63] consumer law, or tort law, other decisions are made on the basis of that content invoking a violation of the service provider's terms and conditions or community standards. The justification for measures taken by the service provider, for example, the removal of such content is essentially contractual; in signing up to the digital platform, the user agrees to abide by the service provider's terms and conditions. If they upload content in violation of those terms and conditions, the service provider has a legal basis to remove it, namely breach of contract.[64] Service providers not only 'legislate' for these contractual terms and conditions, but also adjudicate and enforce them;[65] moreover, not only might the content of these terms and conditions be unclear, or not easily accessible to users but they are offered – as is the case in most legal relationships in which there is a disequilibrium of bargaining power – on a 'take it or leave it' basis. This 'private', often automated, character of content moderation generates concerns with transparency, impartiality, legitimacy and accountability.[66]

This leads us onto the second set of challenges arising in relation to internet governance and regulation, namely the relationship between this private regulatory framework, on the one hand, and the obligations imposed on

[60]The terms and conditions made public by an online platform may however diverge considerably from their internal policies of content moderation; see Catherine Buni and Soraya Chemaly, 'The Secret Rules of the Internet: The Murky History of Moderation, and How It's Shaping the Future of Free Speech' *The Verge* (13 April 2016) <https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech> accessed 17 November 2024.

[61]See, amongst a vast body of literature, Michael P Vandenbergh, 'The Private Life of Public Law' (2005) 105 *Columbia Law Review* 2029, 2037 and Pasquale (n 52) 163.

[62]In relation for example to requirements of online platforms to promote transparency in content moderation concerning terrorist content, see Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online [2021] OJ L172/79.

[63]Similarly, as regards copyright, see Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [2019] OJ L130/92.

[64]Goanta and Ortolani (n 28) 210.

[65]See for example, Luca Belli, Pedro Augusto Francisco and Nicolo Zingales, 'Law of the Land or Law of the Platform? Beware of the Privatisation of Regulation and Police' in Luca Belli and Nicolo Zingales (eds), *How Platforms are Regulated and How They Regulate Us* (FGV Direito Rio 2017) 41, 45.

[66]Giovanni De Gregorio, 'Democratising Online Content Moderation: A Constitutional Framework' (2020) 36 *Computer Law and Security Review* 1, 3.

service providers by State or supranational legislation, which for the purposes of this paper, entails a consideration of the DSA and OSA. This must – it is asserted – be understood more broadly as a reflection of the compatibility of public and private interests.[67] This is the case as the DSA and OSA are responses to the recognition on the part of national governments and Member States, the European Commission and civil society actors, of the limited impact of private actors' self-regulation endeavours.[68]

An outline of the regulatory regimes established at the EU and UK levels

The objectives and scope of the OSA and DSA

The OSA and DSA have broadly common objectives in establishing rules for internet governance, to the ends of making the digital environment safer for those who access it. With the aim of 'making the use of internet services regulated by th[e] Act safer for individuals in the United Kingdom',[69] the OSA establishes duties on service providers 'to identify, mitigate and manage the risks of harm … ',[70] thus rendering companies that provide these services legally responsible for keeping people, and particularly children,[71] safe online. The DSA arguably has a dual objective in the sense that – like many pieces of EU legislation – while it aims to provide for rules promoting a 'safe, predictable and trusted online environment that facilitates innovation and in which fundamental rights enshrined in the Charter, including the principle of consumer protection, are effectively protected', it also aims to facilitate the functioning of the internal market.[72]

A 'safety by design' approach is identifiable in both the OSA and DSA. Both the OSA and DSA establish obligations on all service providers falling within their scope and additional obligations for service providers that satisfy certain conditions, to be balanced with fundamental rights protection. These obligations establish requirements on service providers to undertake risk assessments, provide a framework for content moderation, including setting out transparency obligations. There are some key

[67] José van Dijck, 'The Platform Society as a Contested Concept' in José van Dijck *et al* (eds), *The Platform Society: Public Values in a Connective World* (OUP 2018) 7, 30. See also, Stephanie Law *et al*, 'Private Law, Private Litigation and Public Interest Litigation' (2024) 13 *Global Constitutionalism* 1.
[68] See Benjamin Farrand, 'How Do We Understand Online Harms? The Impact of Conceptual Divides on Regulatory Divergence Between the Online Safety Act and Digital Services Act' (2024) 16 *Journal of Media Law* 1, 3–6.
[69] OSA, s 1(1).
[70] OSA, s 1(2).
[71] Indeed, OSA, s 2(3), aims to provide that children are afforded a higher level of protection than adults.
[72] DSA, Recital 9 and Art 1(1). Indeed, the DSA find its competency basis in Art 114 TFEU, for the approximation of laws for the promotion of the internal market.

divergences as regards the scope of the OSA and DSA depending on the services offered by service providers and the nature of content (and relatedly, the relevant concept of harm, discussed further below).

The OSA establishes duties on service providers that allow users to share content online including social media platforms, and online search engines in relation to regulated[73] 'user-to-user services'[74] and 'search services'.[75] The companies on which such obligations are imposed includes companies that provide services across a breadth of websites, social media sites and apps.[76] In terms of the obligations imposed, the OSA distinguishes between all service providers, and additional obligations on Category 1[77] and 2 services, and those likely to be accessed by children. The OSA itself provides for no criteria on defining these services; instead, it will be defined in secondary legislation.

The DSA may on its face appear to have a broader scope that the OSA. It applies to all 'intermediary services'[78] 'offered to recipients of the service' (that is, essentially service providers offering intermediary services to users); its territorial scope is based on the recipients of the service being established or located in an EU Member State, and not on the place of establishment of the service provider.[79] Thus, it will apply to service providers established outside of the EU. Intermediary services include online platforms (hosting services, which would include social media platforms) and online search engines.[80] Like the OSA, the DSA also distinguishes between service providers depending on their size (and thus, it is anticipated, societal impact), and establishes additional obligations on providers of 'very large online platforms' (hereinafter VLOPs) and 'very large online search engines' (hereinafter VLOSEs).[81]

[73]The notion of 'regulated' is set out in s 4; it essentially rests on the links between the service and the United Kingdom, either if the service has a 'significant number of United Kingdom users', or 'United Kingdom users form one of the target markets for the service' (s 4(5)), or the service can be used in the UK, and may generate a material risk of harm to UK individuals (s 4(6)). As such, the OSA applies to service providers established outside of the UK.

[74]OSA, s 3(1).

[75]OSA, s 3(4). Throughout, this paper refers largely to the legislative provisions concerning user-to-user services; many of the provisions concerning search services are repetitive and the paper aims to avoid this repetition.

[76]This would also include pornographic content services.

[77]Category 1 service providers will include those having the highest risk and impact, user-to-user services; category 2a, those which are broad-reaching search services and category 2b, those which have potentially risky functionalities.

[78]Intermediate services are defined at DSA, Art 3(g) and include 'mere conduit' services, 'caching services' and 'hosting services'. See also DSA, recital 29.

[79]DSA, Art 1(1).

[80]Per DSA Art 3(i) and (j).

[81]These platforms are defined in Art 33 as those with 45 million or more 'average monthly active recipients of service in the Union'. It is acknowledged at recital 76 that these very large platforms must satisfy the 'highest standard of due diligence obligations, proportionate to their societal impact'.

The concept of harm in the OSA and DSA

Fundamentally, a key divergence arises from the concepts of harm engaged in the OSA and DSA. In a shift from the proposal,[82] the OSA establishes duties of care in respect of services where risks of harm and illegality arise;[83] the Act, as it received Royal Assent, governs two categories of harm, namely, 'illegal content and activity' and 'content and activity that is harmful to children'.[84] Per section 234(2), harm is to be understood as 'physical or psychological harm'; further provisions in section 234 entail – as Farrand notes – that the OSA is concerned with harm encountered by individuals, thus excluding broader societal or systemic harms from the scope of the legislation.[85]

In contrast, as Farrand examines in detail,[86] the DSA adopts a concept of harm that engages not only illegality but refers to 'unlawful or otherwise harmful information and activities'[87] and 'illegal or otherwise harmful content and activities'.[88] Moreover, the concept of harm in the DSA engages not only harm or the risk of harm to the individual but to 'societal and economic harm', particularly that which may arise in relation to the use of VLOPs and VLOSEs, where it is noted that the way in which services are designed 'is generally optimised to benefit their own advertising-driven business models and can cause societal concerns'.[89] At recitals 80 to 83, four categories of systemic risk are identified and it is noted that the entire design of VLOPs and VLOSEs, including the use of algorithms, may contribute to these risks.[90] The risks encompass those arising from the spread of illegal content (including child sexual abuse material or illegal hate speech) and unlawful conduct (including the sale of products of services which is prohibited by EU or national law),[91] threats to fundamental rights (particularly those protected by the Charter of Fundamental Rights, and including risks arising from algorithmic systems used by service providers, including

[82]Which originally provided for three categories of harm, including so-called 'legal yet harmful', content that is harmful to adults; see, 'Online Harms White Paper' (n 17), 5–6.
[83]See OFCOM, *Illegal Harms Consultation* (OFCOM, 2023); the consultation elaborates further on the types of services that fall within the scope of the OSA.
[84]OSA, s 1(2)(a) and ss 60 to 63.
[85]Farrand (n 68) 15.
[86]Farrand, ibid.
[87]DSA, Recital 5 referring to information society services, defined in Directive (EU) 2015/1535 of the European Parliament and of the Council of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services [2015] OJ L241/1.
[88]DSA, Recital 68 in the context of advertising.
[89]DSA, Recital 79. The dangers of targeted advertising, particularly their scope to 'negatively impact entire groups and amplify societal harms' is highlighted at DSA, Recital 69. Arts 26 and 27 also provide for further rules on advertising online and the use of recommender systems; these will not be explored further in this paper.
[90]DSA, Recital 84.
[91]DSA, Recital 80.

for example, that of discrimination),[92] 'actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes, as well as public security',[93] and risks arising from the 'design, functioning or use, including through manipulation' of VLOPs and VLOSEs, including impacts on public health, minors, negative effects on physical and mental health, gender-based violence, and coordinated disinformation campaigns.[94]

In comparison to the OSA, the scope of the DSA is wider, both in terms of the service providers to which the rules apply, and as regards the concept of harm. Under the OSA, 'illegal content' is 'content that amounts to a relevant offence';[95] that content may consist of 'words, images, speech or sound'. It will amount to a 'relevant offence' if, the use of such content, possessing, viewing or accessing such content or publishing or disseminating such content, amounts to an offence.[96] What constitutes a 'relevant offence' for the OSA includes existing criminal offences and new offences introduced under the legislation.[97] 'Relevant offences' are sub-categorised into 'priority offences'[98] and 'an offence within subsection 5'.[99] Priority offences encompass terrorism offences,[100] offences related to child sexual exploitation and abuse,[101] and 'other priority offences' per Schedule 7; Schedule 7 includes a range of other offences include assisting suicide, threatening to skill, and disclosing, or threatening to disclose, intimate photographs or videos. Content that amounts to an offence specified in these schedules will constitute priority illegal content.[102] Offences 'within subsection 5' include offences which are not priority offences, in respect of which the victim or intended victim is an individual (or individuals), created either by the OSA itself, or by another piece of legislation.[103] The OSA therefore purports to clearly and comprehensively establish what constitutes illegal content and priority illegal content.

In contrast to the OSA's very precise definition of what constitutes illegal content, and exhaustive listing of 'priority illegal content' the DSA is more open-ended. Article 3(h), DSA encompasses a broadbrush reference to 'any information that, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with

[92]DSA, Recital 81.
[93]DSA, Recital 82.
[94]DSA, Recital 83.
[95]OSA, s 59(1).
[96]OSA, s 59(3).
[97]Offences introduced in the OSA include 'communications offences' under OSA, Part 10.
[98]OSA, s 59(7), and schedules 5 to 7.
[99]Which exclude, for example, violations of IP law and certain violations of consumer law (ie offences under the Consumer Protection from Unfair Trading Regulations 2008, SI 2008/1277) per OSA, s 59(6).
[100]OSA, Schedule 5.
[101]OSA, Schedule 6.
[102]OSA, s 59(8) to (10).
[103]OSA, s 59(5). It is worth noting that offences concerning violations of IP law and certain violations of consumer law do not constitute 'schedule 5' offences, per s 59(6).

Union law or the law of any Member State … ;' it is to be interpreted broadly, and essentially is used to make illegal online that which is illegal offline, and includes for example, 'illegal hate speech, terrorist content and unlawful discriminatory content' or that concerns illegal activities, including the sharing of child sexual abuse, the sale of products or provision of services in violation of consumer law,[104] non-consensual sharing of private images, amongst others.[105]

The obligations on service providers under the OSA and DSA

A 'differentiated duty of care' approach under the OSA

By virtue of section 2(3), OSA the duties on services providers aim to ensure that regulated services are 'safe by design'; services must be offered in a way that mitigates the risk that they are used to make illegal content and activities available to users, while balancing the protection of freedom of expression and privacy. The OSA also establishes obligations on service providers relating to transparency and accountability, to the ends of ensuring that providers are transparent about the content to which users have access, why decisions are made about certain types of content, and to empower users to exercise greater control over the content made available to them.

The OSA adopts a 'differentiated duty of care' approach. Section 7 outlines the duties of care imposed on service providers, further elaborated upon in subsequent provisions of the Act. Section 7(2) establishes duties of care on all providers of regulated user-to-user services; these obligations are differentiated depending on whether the content is illegal or harmful to children (to be determined by the service provider itself), and include a duty to undertake illegal content risk assessment[106] and children's risk assessments where content is likely to be accessed by children.[107] Illegal content risk assessments must take into account the risk profile of the service, and identify the risk of users encountering priority illegal content 'with each kind separately assessed', and other illegal content.[108] Under the OSA, service providers must undertake and keep up to date 'suitable and sufficient' illegal content risk assessments.[109] The assessment must provide

[104]To this end, DSA, Art 30 also sets out that where consumers can enter into distance contracts with traders, the online platforms should provide for mechanisms that allow traders to be traced. See also, Art 31, which broadly requires that online platforms design and operate interfaces in such a way that enables traders to comply with EU consumer law. Per Art 32, consumers who have purchased illegal products or services through its platform should also be informed by the online platform providers of such a violation of EU consumer law.
[105]DSA, Recital 12.
[106]OSA, s 7(2)(a) in line with s 9.
[107]OSA, s 7(4) in line with s 11.
[108]OSA, s 26(5).
[109]OSA, s 9(2) to (3).

a risk profile relating to the service, encompassing the user base, the risk of harm to service users (as regards the risk that illegal content will be presented to them, and that those services may be used for the commission of offences), the nature and severity of that harm, and the way in which the service's functionalities, including use of algorithms, may lead to the spread of illegal content.[110]

In relation to illegal content (and similarly as regards content that is harmful to children)[111], the OSA establishes duties on all service providers[112] in relation to the service as a whole, including its design and operation, regulatory compliance and risk management, use of algorithms, terms of service, and content moderation, amongst others.[113] It is here that one can find reference to the limited rules specifically on content moderation in the OSA. Service providers must essentially evaluate their risks assessments and take proportionate measures to preclude individuals from accessing or being confronted with illegal content, mitigate and manage the risk of the service being used for illegal purposes, and mitigate and manage the risk of harm to individuals in light of their illegal content risk assessment.[114] Service providers should equally take proportionate measures to ensure that illegal content is present for a minimal length of time, and that, if alerted to the presence of such content or otherwise becomes aware of its presence, ensure it is removed 'swiftly'.[115] The terms and conditions of the service provider must be 'clear and accessible', identify how they are protecting individuals from illegal content, including any use of automation; these terms must also be applied consistently.[116] Service providers may become aware of illegal content through risk assessments, or because it has been flagged to them. Section 20 requires that services providers establish mechanisms that allow 'affected persons' to flag and report content that is illegal, or harmful to children.[117] Service providers should also establish a procedure which allows users and affected persons to bring complaints in relation to illegal or content that is harmful to children, complaints that duties of care are not being complied with, complaints by users in respect of a decision to remove content on the basis of illegality or other measures taken (including, for example, suspension of an account), and complaints in relating to the use of automated content moderation.[118]

[110]OSA, s 9(5).
[111]OSA, s 12 in relation to user-to-user services; s 29 in relation to search services.
[112]s 7(2) in line with s 10, in relation to user-to-user services and s 27 in relation to search services. While establishing an additional reporting obligation on most recent illegal content risk assessments in s 10(9) for 'category 1' services.
[113]OSA, s 10(4).
[114]OSA, s 10(2).
[115]OSA, s 10(3).
[116]As set out at OSA, s 10(5) to (9).
[117]'Affected persons' encompass the subject of the content but may also include third parties, 'a member of a class or group of people with a certain characteristic targeted by the content', per OSA, s 20(4).

Engaging the distinction between priority illegal content (concerning priority offences) and other illegal content, the OSA thus establishes different safety duties depending on the category of content. For illegal and priority illegal content, service providers are obliged to take or use proportionate measures relating to the design or operation of the service to effectively mitigate and manage the risk of harm to individuals arising from that content, and – when they are alerted to the presence of such content on the service – or when they become aware of it, they must use proportionate systems and processes to swiftly remove it. For priority illegal content, additional obligations are established. Service providers must establish and utilise proportionate measures as pertains to the service's design or operation in order to: prevent individuals from encountering priority illegal content though that service;[119] mitigate and manage the risk of the service being used to commit or facilitate a priority offence (as identified by virtue of the risk assessment outline above);[120] and minimise the length of time priority illegal content is present on the service.[121] Moreover, service providers must provide for clear and accessible information in their terms and conditions as to how users are to be protected from illegal content, and make separate provision as regards protection from priority illegal content.[122]

In terms of the nature of the regulatory and enforcement regime envisaged by the OSA, the regime has been characterised as 'co-regulatory' in its nature as OFCOM[123] plays a key role in the oversight of service providers' compliance with their duties of care and in the enforcement of the OSA.[124] The establishment of duties of care on service providers entails that service providers themselves have a key role to play in content moderation; that is to say, in identifying the risk of harm arising from illegal content and content that is harmful to children, in mitigating the risks that users will encounter such material and in removing that content, when they become aware of its presence on their platforms. Section 192 requires that service providers remove content, where, using 'reasonably available' information, there are 'reasonable grounds to infer' it is illegal.

[118]OSA, s 21.
[119]Including under OSA, s 27, an obligation to 'operate a service using proportionate systems and processes designed to minimise the risk of individuals encountering search content' which is priority illegal content, or other illegal content (where, as regards the latter, the service provider has been alerted to it, or become aware of it).
[120]OSA, s 10(2).
[121]OSA, s 10(3).
[122]OSA, s 10(5).
[123]The OSA empowers OFCOM as the independent regulatory body for online safety. It is empowered with drafting guidance for service providers and implementing codes of practice to facilitate their compliance with the OSA, as well as monitoring, investigatory and enforcement powers. These are explored further below.
[124]See for example, Farrand (n 68) 19.

Like the DSA, the OSA does not require service providers to engage the broad monitoring of content (with the exception of priority illegal content, as set out above, per the safety duties in section 10(2) and (3), OSA), nor does it impose liability for the storage of information. Rather, service providers must comply with the duties of care applicable to them. However, the OSA itself provides for very little in terms of content moderation in relation to content that is not illegal (other in relation to content that poses a risk to minors, ie content that is harmful to children), including, for example, disinformation. This is in contrast to the DSA, which sets out a much more comprehensive framework in relation to content moderation, with additional obligations on VLOPs and VLOSEs, as examined below. In relation to category 1 service providers, in addition to risk assessments concerning illegal content, the OSA requires that 'user empowerment' tools are offered in relation to 'adult user content'. To this end, sections 14 to 16 entail that category 1 service providers must provide mechanisms that offer users greater control over harmful material that they might encounter. Such harmful material would include, for example, content on suicide, or self-harm, abusive content, for example concerning race, religion, sex, sexual orientation and gender reassignment, or content that incites hatred.[125] Service providers are required to undertake, maintain and record assessments for the purposes of user empowerment obligations.[126] The mechanisms to be provided by service providers must be easily accessible, proportionate and operate to reduce the risk of such content being encountered or alert the user to the presence of such content.[127] This may entail, depending on the risk assessment undertaken, that the service provider take proactive measures to remove this content.

At present, gaps remain in the OSA in relation to content moderation, particularly where it goes beyond dealing with illegal content, that is, where decisions are made and measures are taken on the basis of the service providers' terms and conditions. The way in which these obligations will be imposed will be determined with future guidance, codes of practice[128] and secondary legislation.[129] The absence of specific provisions in the OSA

[125]OSA, ss 15(3) to (5).
[126]OSA, ss 14(3) and (4).
[127]OSA, ss 15(3) and (4).
[128]Moreover, Phase Three of OFCOM's consultations closed in May 2023; OFCOM, 'Third Phase of Online Safety Regulation: Additional Duties for Categorised Services, Call for Evidence' 25 March 2024 <https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-2-6-weeks/281475-third-phase-of-online-safety-regulation/associated-documents/online-safety-phase-3-call-for-evidence.pdf> accessed 12 June 2024. OFCOM is obliged to publish codes of conduct and guidance for regulated services. See also, OFCOM, 'Content Moderation in User-to-User Online Services: An Overview of Processes and Challenges' 14 September 2023.
[129]The UK's Information Commissioner's Office has published a guidance document on content moderation and data protection obligations, which is also relevant to content moderation undertaken by service providers: ICO, 'Content Moderation and Data Protection', February 2024 <https://ico.org.uk/

on content moderation seem to stem from its focus on priority or other illegal content (and content that is harmful to children) in particular; this focus dictates that any competence under the OSA to require service providers to deal via content moderation with priority or other illegal content (or content harmful to children) is necessarily limited.[130] This is in stark contrast to the DSA, which – as is set out below – makes comprehensive provision in relation to content provision, including decisions made and measures taken by service providers on the basis of their terms and conditions, as well as on measures taken in relation to illegal content.

Due diligence obligations under the DSA

The DSA establishes due diligence obligations on service providers, alongside obligations of fairness, transparency and accountability in relation to their content moderation processes. Like the e-Commerce Directive, the framework of the DSA is based on the notion that service providers[131] are not liable for the storage of information, on the condition that they have no 'actual knowledge of illegal activity or illegal content' or 'upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the illegal content'.[132] In order to promote legal certainty in content moderation,[133] that is, activities by which service providers 'detect, identify and act against illegal content',[134] these exceptions from liability are not lost through service providers' 'voluntary own-initiative investigations', by which they aim to identify and remove illegal content to the ends of complying with Union or national law.[135] Moreover, service providers are under no obligation to monitor or engage in fact-finding in relation to content.[136] The exemption from liability is however conditional; this means that service providers must, if they actually become aware of illegal activities or content, act quickly to remove it or disable access to it. This should be done in line with the fundamental rights, including those of

media/for-organisations/uk-gdpr-guidance-and-resources/online-safety-and-data-protection/content-moderation-and-data-protection-0-0.pdf> accessed: 24 June 2024.

[130]See for example, Lorna Woods *et al*, 'Strengthening the user empowerment tools in The Online Safety Bill' Carnegie Trust UK, 6 July 2023 <https://carnegieuktrust.org.uk/blog-posts/strengthening-the-user-empowerment-tools-in-the-online-safety-bill/> accessed 24 June 2024.

[131]For the purposes of this paper, hosting services.

[132]DSA, Art 6(1)(b), alongside Arts 4 and 5 in relation to conduit and caching services.

[133]Content moderation is defined at Article 3(t), DSA to mean … means the activities, whether automated or not, undertaken by providers of intermediary services, that are aimed, in particular, at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service, including measures taken that affect the availability, visibility, and accessibility of that illegal content or that information, such as demotion, demonetisation, disabling of access to, or removal thereof, or that affect the ability of the recipients of the service to provide that information, such as the termination or suspension of a recipient's account.

[134]DSA, Recital 26.

[135]DSA, Art 7.

[136]DSA, Art 8.

freedom of expression and of information, of the service recipients.[137] Recital 12 further provides that the service provider may become aware of this illegality through its own investigations or on notice of an individual or entity.[138]

Like the OSA, the DSA adopts an approach of co-regulation; that is to say, the DSA envisages a role to be played by service providers in content moderation (which might be deemed to reflect a form of self-regulation) and establishes obligations on service providers within the legislation itself, providing for mechanisms for national authorities and the European Commission in implementing these obligations and coordinating their activities. To the ends of facilitating content moderation, on the one hand, and the enforcement of the DSA, on the other, service providers must provide contact points, both for Member States' authorities (Digital Service Coordinators, DSC), the Commission and the European Board for Digital Services (hereinafter EBDS),[139] as well as for service users.[140] Generally, service providers do engage in these 'own-initiative investigations' and content moderation, which often entails the use of algorithms or automated detection of illegal content.[141]

The DSA does not preclude service providers from removing or limiting access to content that – while not illegal – is in contravention of their terms and conditions. This is done by virtue of contract, that is to say, the user's agreement to the provider's terms and conditions in which certain content or services is prohibited. As private entities, service providers are generally unrestricted under the DSA in terms of what their terms and conditions can provide for, as long as those terms and conditions are not unlawful;[142] nevertheless, it establishes transparency rules in relation to the accessibility of terms and conditions.

Article 14 establishes that service providers' terms and conditions must be publicly available, and set out any restrictions established therein as regards the use of their services; the content of terms and conditions must be easily accessible to users (including children), and expressed in 'clear, plain,

[137]DSA, Recital 22.

[138]DSA, Recital 22 – 'in so far as such notices are sufficiently precise and adequately substantiated to allow a diligent economic operator to reasonably identify, assess and, where appropriate, act against the allegedly illegal content.'

[139]DSA, Art 11. The EBDS, its structure and tasks are set out at Arts 61 to 63.

[140]DSA, Art 12.

[141]See Ortolani (n 34) fn 35.

[142]See Ortolani, ibid 6. Ortolani also discusses the notion that, given these private entities – notwithstanding that they may operate to provide for what might be deemed 'public' services, or services that States might provide, and have a significant societal impact and role – are private, they are not bound in the same way as States are, to human rights obligations, including those relating to freedom of expression; Ortolani notes however, that Meta's Oversight Board, will often evaluate Meta's terms and conditions in light of human rights (referring to Lorenzo Gradoni, 'Chasing Global Legal Particles: Some Guesswork about the Nature of Meta's Oversight Board', EJIL: Talk! <https://www.ejiltalk.org/chasing-global-legal-particles-some-guesswork-about-the-nature-of-metas-oversight-board/> accessed 14 June 2024.

intelligible, user-friendly and unambiguous language'[143] while any 'significant changes' must be communicated.[144] This also entails that information must be provided on the processes and purposes of content moderation (including policies, procedures, measures and tools), and should indicate if any algorithms are used in these processes, as well as any 'human in the loop' of the decision making and review.[145] Service providers are required to engage in content moderation in a 'diligent, objective and proportionate manner', taking into account the fundamental rights and legitimate interests of parties.[146] In relation to hosting services, Article 16 requires hosting service providers to enable individuals and entities (thus, users and third parties), to notify the provider of illegal content; this should be facilitated by mechanisms that are easily accessible and user-friendly, allow for electronic submissions of 'sufficiently precise and adequately substantiated notices', including explanations as to why there is a belief the content is illegal. Such notices are deemed to constitute the 'actual knowledge or awareness' for the purposes of Article 6.

In terms of third-party engagement, as regards online platform providers, the DSA also establishes a system of 'trusted flaggers' for Article 16 notification purposes. Under Article 22, entities may apply to the national DSC to be a trusted flagger,[147] if they have the 'particular expertise and competence for the purposes of detecting, identifying and notifying illegal content', are independent and submits 'notices diligently, accurately and objectively'.[148] Trusted flaggers must (at least annually) report to the national DSC and make public data indicating the relevant services providers, the type of illegal content and the action taken by the service provider.[149]

The service provider (including online platforms) is required to provide an acknowledgement and a reasoned response (without undue delay) to the poster of the content including the decision taken, alongside information on avenues of redress (whether internal complaint processes, out of court dispute resolution or judicial redress).[150] The decision should be 'clear and easily comprehensible and as precise and specific as reasonably possible',[151]

[143]DSA, Art 14(1).
[144]DSA, Art 14(2).
[145]VLOPs and VLOSEs must also publish their terms and conditions in all EU languages in which they provide services (Article 14(6)) and provide users with 'a concise, easily accessible and machine readable summary of the terms and conditions, including the available remedies and redress mechanisms' (Art 14(5)); this is a reflection of the importance of these terms and conditions and the recognition that users do not tend to read the entirety of lengthy documents.
[146]DSA, Art 14(4).
[147]DSCs must inform the Commission of the trusted flaggers established within their jurisdiction, which is to be published by the Commission per Arts 22(4) and (5). Trusted flagger status can also be suspended or revoked by the DSC under Arts 22(6) and (7).
[148]DSA, Art 22(2).
[149]DSA, Art 22(3).
[150]DSA, Art 17(3)(f).
[151]DSA, Art 17(3)(4).

reasoned and issued to 'any affected recipients of the service' of any restrictions,[152] including the removal or limitation of access to content or services, and the basis on which it has been removed (whether it is illegal or a prohibition deriving from the terms and conditions).[153] Decisions may provide that the content should (or should not) be removed, that access to that content should (or should not) be disabled, that its visibility should (or should not) be restricted, that the provision of the service should be suspended or terminated (or not), that the recipient's account should be suspended or terminated (or not), or that the recipient's ability to monetise information should be suspended, terminated, or restricted (or not).[154] It is worth noting that service providers may play a role in relation to criminal law enforcement; Article 18 requires hosting service providers to inform national law enforcement agencies of any suspicions of criminal offences, including those which generate a threat to an individual's life or safety.[155]

Additional obligations of VLOPs and VLOSEs under the DSA

The DSA establishes additional obligation on platforms that are designated as VLOPs and VLOSEs by the Commission, following consultation with Member States' DSCs, in accordance with Article 33.[156] VLOPs and VLOSEs are obliged to undertake 'specific … and proportionate' risk assessments, at least annually,[157] in order to 'diligently identify, analyse and assess any systemic risks … stemming from the design or functioning of their service and its related systems, including algorithmic systems, or from the use made of their services'.[158] The risks of harm anticipated under Article 34 are broader than under the OSA. Already in Article 34(1), the DSA refers to 'systemic risk'; this entails an assessment of service providers' content moderation process, terms and conditions, use of algorithms, advertising and recommender systems and use of data,[159] as well as risks of illegal content, 'actual or foreseeable negative effects' on fundamental rights, civic discourse, electoral processes and public security, gender-based violence, public health protections and physical and psychological well-being.[160] VLOPs and VLOSEs must also establish 'compliance functions' to monitor their compliance with the DSA, including their Chapter III obligations;[161]

[152]DSA, Arts 17(1)(a) to (d).
[153]DSA, Art 17(3).
[154]DSA, Arts 17(1)(a) to (d).
[155]DSA, Art 18.
[156]These platforms are defined in DSA, Art 33, as those with 45 million or more 'average monthly active recipients of service in the Union'. It is acknowledged at Recital 76 that these very large platforms must satisfy the 'highest standard of due diligence obligations, proportionate to their societal impact'.
[157]Maintaining records for at least three years, per DSA, Art 34(3).
[158]DSA, Art 34(1).
[159]DSA, Art 34(2).
[160]DSA, Arts 34(1)(a) to (d).

compliance is also evaluated via yearly independent audits[162] and, to the ends of transparency, reporting obligations.[163]

In relation to these risks, VLOPs and VLOSEs are also required to 'put in place reasonable, proportionate and effective mitigation measures', which may entail making changes to their internal processes, including content moderation, taking targeted measures (for example, in respect of children), amending their terms and conditions, use of recommender systems, as well as the design and operation of their services.[164] Both the Commission and Member States' DSCs have a role to play in relation to the identification of such systemic risks and the platforms' mitigation strategies. In response to the submission of risk assessments, the Commission and national DSCs, can *inter alia* provide guidance on best practices and measures to be taken by service providers.[165] Where a risk of serious threats to public security or health are identified, the Commission, on the recommendation of the EBDS, can initiate a 'crisis response mechanism' on the basis of Article 36 and Article 48.

Internal complaint-handling system and out-of-court dispute settlement

In contrast to the OSA, which makes limited provisions on service providers' complaint procedures, the rules established in the DSA are more extensive and comprehensive. Online platform providers are also required to establish a 'timely, non-discriminatory, diligent and non-arbitrary'[166] internal complaint-handling system in line with Article 20. This system should be open to recipients of the service, users and third parties,[167] for at least six months following the service provider's decision on an Article 16 notice.[168] It should be effective, user-friendly, free and allow for the submission of 'sufficiently precise and adequately substantiated' complaints electronically,[169] in respect of decisions taken by service providers, outlined above. The service provider should reverse its decision to take or not take action if the complaint provides 'sufficient grounds' for such an outcome; a reasoned decision should be given, and the service provider should inform the individual of scope for further dispute resolution processes, including the Article 21 out-of-court dispute settlement procedure[170] and

[161]As well compliance with any codes of conduct with which they have voluntary complied under DSA, Arts 45 to 47, DSA.
[162]DSA, Art 37.
[163]DSA, Art 42.
[164]DSA, Art 35(1).
[165]DSA, Art 35(2).
[166]DSA, Art 20(4).
[167]DSA, Art 20(1) refers to 'recipients of the service, including individuals or entities'.
[168]DSA, Art 20(2).
[169]DSA, Art 20(3).

judicial redress.[171] Importantly, this internal complaints system should not encompass only automated decision making; there must be an 'appropriately qualified'[172] human 'in the loop'. This entails a degree of expertise.

The DSA also anticipates the use of alternative dispute resolution-type mechanisms to further deal with the resolution of disputes following the submission of an Article 16 notice; this includes complaints not resolved by the internal process set about in Article 20, and is open to recipients of the service, users and third parties. Like consumer ADR platforms, online platform providers must provide information about out-of-court settlement processes to its potential users. Dispute settlement bodies must be certified by the DSCof the Member State in which it is established.[173] Bodies must be capable of facilitating the 'swift, efficient and cost-effective' settlement of disputes;[174] that is to say, they must be impartial, independent (also financially) and have the 'necessary expertise' concerning illegal content and the terms and conditions of at least one type of online platform. Dispute settlement processes must be easily accessible and allow for procedural steps to be completed online.[175] The use of the out-of-court settlement process does not preclude access to the courts; indeed, the decision of the out-of-court settlement body is not binding.[176]

As above, service providers are not required to engage in general monitoring; however, in order to avoid misuse of their services and content moderation processes,[177] online platform providers are required to keep records of both the frequency by which service recipients provide illegal content, and the frequency by which individuals or entities submit manifestly unfounded notices or complaints. Article 23 requires that in relation to the former, the provision of services, and in relation to the latter, the processing of notices and complaints, shall be, following a warning, suspended for a reasonable

[170]An example of a recently certified out of court dispute settlement body under DSA, Art 21 is the Appeals Centre Europe, established by the Irish Digital Services Coordinator, Coimisiún na Meán. For an initial analysis, see Pietro Ortolani and Lorenzo Gradoni, 'Op-Ed: ACE Up the Sleeve? Appeals Centre Europe (ACE) as Meta's Bid to Hegemonise Social Media Content Moderation in the EU?' *EU Law Live* <https://eulawlive.com/op-ed-ace-up-the-sleeve-appeals-centre-europe-ace-as-metas-bid-to-hegemonise-social-media-content-moderation-in-the-eu/> accessed 18 November 2024.

[171]DSA, Arts 20(4) and (5).

[172]DSA, Art 20(6).

[173]DSA, Art 21(3). They also have reporting obligations, on an annual and biannual basis (reporting on their functioning, numbers of disputes received, outcomes, timing for resolution and challenges arising, and in relation to the latter, also on systemic challenges, best practices and recommendations for improvement) per DSA, Art 21(4). The process should be free for service recipients, users and other entities, or incur only a nominal fee; fees should be clearly communicated, and reasonable. If the body decides against the service provider, the service provider incurs all fees; if the body decides in its favour, fees are only charged if it is found that the party bringing the dispute acts in bad faith, per DSA, Art 21(5).

[174]Decisions must be taken within a 90-day period, or a maximum of 190 days, if the dispute is highly complex per DSA, Art 21(4).

[175]DSA, Art 21(3).

[176]DSA, Art 21(2).

[177]Their policy on misuse should also be set out in their terms and conditions per DSA, Art 23(4).

time period. A decision to make such a suspension is to be made on an individual basis, and take into account all relevant facts and circumstances, including the number and proportion of pieces of illegal content and manifestly unfounded notices or complaints, the gravity of the misuse, and the intentions of the parties involved.[178]

Reporting obligations under the DSA

Similarly to the ends of transparency, service providers must report (at least on a yearly basis), and make public, information on content moderation during that period; this includes an indication of Article 9 and 10 orders,[179] as well as 'meaningful and comprehensible information' concerning voluntary content moderation processes, the use of automated tools,[180] training on content moderation, the number and types of measures taken and the types of illegal content involved or violations of the terms and conditions concerned.[181] Online platform providers must report on the number of Article 16 notices, the types of illegal content to which they relate, the number of notices submitted by trusted flaggers, action taken (and the basis of such action, whether it concerned illegality or the provider's terms and conditions), the use of automated tools and the average time for the process.[182] The service provider must also report on complaints received and resolved in line with their terms and conditions via their internal complaints process, while online platform providers must indicate, per Article 20, also the basis for such complaints and decisions taken in respect of them, the average time for the process and an indication of the number of times decisions are reversed.[183]

Article 24 elaborates on these transparency obligations for online platform providers in relation to the use of the Article 21 out-of-court settlement procedure[184] and Article 23 decisions on suspensions in relation to misuse.[185] Twice a year, online platforms are also required to report on the 'average monthly active recipients of the service in the Union', and must provide such information to its national DSC or the Commission if requested.[186] This data is used to determine if the service provider constitutes a VLOP or VLOSE in accordance with Article 33.

[178]DSA, Art 23.
[179]DSA, Art 15(1)(a).
[180]DSA, Art 15(1)(e).
[181]DSA, Art 15(1)(c).
[182]DSA, Art 15(1)(b).
[183]DSA, Art 15(1)(d).
[184]DSA, Art 24(1)(a).
[185]DSA, Art 24(1)(b).
[186]DSA, Arts 24(2) and (3).

The enforcement regimes of the OSA and DSA

The role of OFCOM under the OSA: a single (national) regulator

The powers and duties of OFCOM

The powers and duties of OFCOM are set out at Part 7, OSA.[187] OFCOM has a role to play in monitoring service providers; it firstly must evaluate service providers and maintain and publish a register of regulated services.[188] OFCOM is also obliged to undertake risk assessments of regulated services in light of their characteristics[189] to identify the risks of harm to individuals (in relation to illegal content and content that is harmful to children);[190] the outcomes of these risk assessments and risk profiles of services providers should be published.[191] OFCOM must provide also guidance and draft codes of conduct for service providers in order to facilitate their compliance with duties of care concerning illegal content and content that is harmful to children; this includes, for example, guidance for service providers on completing risk assessments to facilitate their compliance with their duties to provide such risk assessments.[192] OFCOM is engaged in a consultation process with stakeholders, which began at the end of 2023, and is likely to continue until mid-2025. Moreover, in respect of the enforcement framework under the OSA, OFCOM has been attributed monitoring, investigatory and enforcement powers,[193] detailed in Part 7, Chapter 4.

OFCOM is empowered under sections 100 to 104,[194] OSA to require service providers to furnish information, in a way that is proportionate to OFCOM's exercise of its functions,[195] to the ends of monitoring service providers' compliance with their duties of care under the OSA.[196] In order to obtain such information, OFCOM must issue an information notice in line with section 100(1) OSA, setting out the information that they require and why, the manner in which information should be provided and consequences of non-compliance with such a notice.[197] Failure to comply with information notices is deemed to constitute an offence under section 109,

[187]Amending the duties of OFCOM under the Communications Act 2003.
[188]OSA, ss 95(2) and (3), s 95(8). NB: the OSA distinguishes between category 1 (regulated user-to-user services in s 97), 2A (regulated search services and combined services) and 2B services (regulated user-to-user services) depending on the nature of the services offered (see s 95(10), OSA).
[189]OSA, s 98. Characteristics of service providers are to be understood as including 'its functionalities, user base, business model, governance and other systems and processes' per OSA, s 98(11).
[190]OSA, s 98.
[191]OSA, ss 98(4) and (7).
[192]OSA, s 99.
[193]OSA, s 1(2)(b).
[194]Section 100 establishes the general power of OFCOM to require service providers to provide information, while section 101, for example, requires service providers to provide information in specific circumstances, namely in the instance of a death of a child.
[195]OSA, s 100(4).
[196]OSA, s 100(6).
[197]OSA, s 102(3). In line with the obligation on the person issued with this information notice to comply per OSA, s 102(8).

OSA.[198] Where OFCOM considers it necessary for the purposes of understanding the risk of non-compliance with obligations under the DSA and mitigating the risks of such harm,[199] or where OFCOM considers a service provider to have failed, or be failing, in satisfying their duties of care under the OSA,[200] OFCOM can require a skilled person to be appointed to provide such information.[201] The service provider must provide assistance to the skilled person in preparing this report.[202]

The OSA also empowers OFCOM with investigatory powers under section 105, with which service providers are obliged to comply. This includes powers to undertake interviews (including with employees of service providers),[203] powers to enter premises and seize materials[204] for the purposes of the investigation and powers to undertake audits.[205] Failure to comply with these powers of OFCOM, for example, by destroying information,[206] or providing false information for audit purposes,[207] may constitute an offence under sections 111 to 112. Such 'information offences' may result in penalties being imposed under section 113.

In situations in which service providers are in non-compliance with their obligations under the OSA, OFCOM is empowered under the provisions of Part 7, Chapter 6 with enforcement powers. OFCOM can issue provisional notices of a contravention of the OSA if there are 'reasonable grounds' for considering that the service provider has failed or is failing in relation to an enforceable duty or duties[208], including for example, their duties of care in relation to illegal content risk assessments and illegal content.[209]

The issuance of such contravention notices may also be issued for a failure or failing, to comply with information notices and skilled persons reports, as set out above.[210] Contravention notices should set out the duty or duties with which OFCOM considers the service provider has failed to comply, and the rationale for that opinion,[211] and set out the steps that should be taken to bring the service provider into compliance or to remedy the failure to comply.[212] The notice must also provide information on any proposed

[198]There are limited defences to such a failure, including that it was not 'reasonably practicable' to comply within the time required, or that 'reasonable steps' were taken to comply, per OSA, s 109(2).
[199]OSA, s 104(1)(b).
[200]OSA, s 104(1)(a).
[201]OSA, s 104(4).
[202]OSA, s 104(7).
[203]OSA, s 106.
[204]OSA, s 108.
[205]OSA, s 107.
[206]OSA, s 111(3).
[207]OSA, s 111(2).
[208]The full list of enforceable duties is set out in OSA, s 131.
[209]OSA, s 130.
[210]OSA, s 130(4).
[211]OSA, s 130(5).
[212]OSA, s 130(6).

penalty, including its rationale, and how the amount should be calculated.[213] The notice should also set out that the service provider can, within a specific period of time, make representations to OFCOM in relation to the failure raised.[214] The duty or requirement in relation to which the provision contravention notice has been issued is known as a 'notified requirement'.[215] Following the issuance of a provisional contravention notice and the expiry of the time allowed for representations by the service provider, OFCOM can either issue a 'confirmation decision' under section 132 or decide not to take any further action (in relation to the latter, OFCOM must inform the service provider).[216] The confirmation decision reflects a determination by OFCOM that the service provider is 'has failed, or is failing, to comply with a notified requirement'.[217]

A section 132 confirmation decision may require[218] the service provider to take steps to comply with the notified requirement[219] or remedy the failure to comply with the notified requirement,[220] pay a penalty under section 137, or require both that steps are taken to bring themselves into compliance and make payment of a penalty. Where OFCOM requires that steps are taken by the service provider to bring themselves into compliance, OFCOM must set out the steps that should be taken, the reasons for OFCOM's decision to require such steps to be taken linking each step to the relevant required notifications, set out the time period in which non-compliance has been observed and whether the failure to comply is continuing, and specify a 'reasonable period' within which such steps must be taken.[221] OFCOM must also inform the service provider of their rights of appeal (under section 168) and of the consequences of their non-compliance with the confirmation decision (including scope for enforcement action).[222] Confirmation decisions may themselves become enforceable, for example, if OFCOM seeks an injunction or other measure in civil proceedings under section 133(9).

The confirmation decision may require the service provider to make payment of a single penalty, or a daily rate penalty in relation to continuous failures. Again, OFCOM must provide reasons for its decision to impose a penalty and its calculation, link the penalty to notified requirements, specify the period of the failure to comply, and notify the service provider

[213]OSA, s 130(7).
[214]OSA, s 130(8).
[215]OSA, s 132(1).
[216]OSA, s 132(2).
[217]OSA, s 132(3).
[218]OSA, s 132(5).
[219]Similarly, OFCOM can issue confirmation injunctions in relation to service providers' failures to comply with risk assessment duties and children's access assessments under ss 134 and 135 respectively.
[220]OSA, s 133.
[221]OSA, s 133(4).
[222]OSA, ss 133(4)(h) and (i).

of their right of appeal and consequences of non-payment.[223] Failure to comply with requirements set out in the confirmation decision may constitute an offence (in particular, if the failure relates to the safety of children online), which may lead – following summary convictions – to the imposition of a penalty and/or imprisonment.[224] If no penalty has been imposed in the confirmation decision itself, a penalty can be imposed for failure to comply with the confirmation decision under section 139 (providing OFCOM has satisfied the requirements of the confirmation decision).[225]

OFCOM has been empowered to take different actions to restrict the business of service providers, if they consider that the provider of a service ancillary to a regulated service[226] continues to fail to comply with an 'enforceable requirement'. Service restriction orders[227] (and interim service restriction orders) can be sought by OFCOM from domestic courts, where the service provider has failed and continues to fail to comply with an enforceable requirement, and where the failure relates to requirements established in a confirmation decision or a penalty imposed.[228] OFCOM may also seek such orders if it can establish that a service provider would be unlikely to comply with the requirements of a confirmation decision is issued, or if it considers the 'circumstances of the failure or the risks of harm to individuals in the UK are such that it is appropriate' to seek the order without the earlier steps having been taken or without waiting to determine if compliance can be made (ie with a contravention notice, or confirmation decision).[229] OFCOM's application must satisfy certain procedural requirements; that is to say, it must identify the relevant service, the service provider, the grounds and evidence for the application, what requirements should be imposed, and the relevant persons on whom they should be imposed and evidence they are providing an ancillary service (and if no notice has been given, why).[230] The court can make such an order providing the grounds in section 144(6) are satisfied; in particular, the restriction of service must be deemed to be 'appropriate' to the ends of preventing harm to UK individuals and proportionate to the risk of harm.[231] The service restriction order essentially has the effect of setting requirements in relation to

[223]OSA, s 137(5).
[224]OSA, s 138.
[225]The amount of penalties is set out in OSA, Schedule 13.
[226]See OSA, s 144(11); 'ancillary' services are those which 'facilitates the provision of the regulated service (or part of it), whether directly or indirectly, or displays or promotes content relating to the regulated service' and include, for example, per s 114(12), 'search engines which generate search results displaying or promoting content relating to a regulated service' or 'user-to-user services which make content relating to a regulated service available to users'.
[227]OSA, s 144.
[228]OSA, s 144(3).
[229]OSA, ss 144(4)(c)(iii) and (iv).
[230]OSA, s 144(5).
[231]OSA, s 144(6).

> withdrawing the ancillary service to the extent that it relates to the relevant service (or part of it), or preventing the ancillary service from promoting or displaying content that relates to the relevant service (or part of it) in any way.[232]

Such orders should set out a time period for compliance and should have an expiry date.[233] OFCOM is also responsible for informing the Secretary of State when such orders are made (within a reasonably practicable period).[234]

Where service restrictions orders are not sufficient to prevent significant harm to UK individuals, OFCOM can also apply to the court for an access restriction order (or interim order),[235] which would impose 'requirements on one or more persons who provide an access facility (whether from within or outside the United Kingdom) in relation to a regulated service', where an access facility is understood to mean a facility provided which can be 'withdraw[n], adapt[ed] or manipulate[d] in such a way to impede access … to the regulated service by United Kingdom users of that service'.[236]

Importantly, for transparency purposes, OFCOM must make public any enforcement action taken, including confirmation notices and penalty notices, indicating the relevant persons, failures and OFCOM's response[237] and can require that the relevant person also make a publication notice.[238] In relation to these enforcement actions and how they purport to exercise their powers, OFCOM must under section 151 provide guidance for service providers.

Appeals and 'super-complaints'

There is scope to appeal in relation to determinations and decisions made by OFCOM. Section 167 sets out the scope for providers to appeal, before the Upper Tribunal, in relation to OFCOM's decisions on categorising service providers under Section 95 (ie OFCOM's register of 'Part 3' services providers, that is, providers of regulated user-to-user services and regulated search services). This is in essence a process of judicial review of OFCOM's categorisation decision.[239] OFCOM's decisions to issue notices, including confirmation decisions and penalty notices may also be appealed under section 168. Such appeals can be brought by those issued with the notice and (where leave of the Upper Tribunal is given) sufficiently interested persons.[240] Again, this is essentially a process of judicial review.[241]

[232]OSA, s 144(8)(c).
[233]OSA, ss 144(d) and (e).
[234]OSA, s 144(10).
[235]OSA, ss 146 and 147.
[236]OSA, s 146(10); this would include per s 146(11), internet access services and app stores.
[237]OSA, s 149.
[238]OSA, s 150.
[239]OSA, s 167(5).
[240]OSA, s 168(2).
[241]OSA, s 168(3).

The OSA also establishes the framework for a 'super-complaints' process, by which an 'eligible entity' can complain to OFCOM in relation to the features or conduct of service providers, where it considers that there is a 'material risk' of (a) significant harm to service users or the public, (b) a significant adverse effect on the freedom of expression of users or the public, or (c) a significant adverse impact on users or the public.[242] The admissibility of such complaints depends on its importance or whether a particularly large number of users or members of the public are affected.[243] The criteria for eligible entities is to be set by the Secretary of State; these are likely to be bodies 'representing the interests of users of regulated services, or members of the public'.[244] The super-complaints procedure is to be established by regulations made by the Secretary of State, which are not set out in the OSA itself, in consultation with OFCOM. OFCOM should similarly draft and publish guidance on the criteria of the super-complaints procedure, including procedural matters.[245]

A multilevel approach to the enforcement of the DSA: the role of the Member States' digital service coordinators and the European Commission

The DSA establishes a novel supervision and enforcement framework in Article 49, albeit one that reflects somewhat the enforcement framework of other pieces of EU legislation, including the GDPR.[246] Member States are required to designate competent authorities, of which one should be designated as the 'Digital Services Coordinator' (DSC),[247] with responsibility for supervising and enforcing the DSA within that Member State and coordinating with other DSCs across the EU.[248]

DSCs should be independent from both public and private influence, issue annual reports on their activities,[249] and have the necessary resources made available to them by their Member States.[250] While service recipients, and 'any body, organisation or association mandated to exercise the rights conferred by this Regulation on their behalf',[251] may bring complaints before the

[242]OSA, s 169(1).
[243]OSA, s 169(2).
[244]OSA, ss 169(3) and (4).
[245]OSA, s 171.
[246]GDPR, Art 51.
[247]These DSCs as well as their contact information, must be communicated to the EBDS and Commission per DSA, Art 49(3).
[248]DSA, Arts 49(1) and (2).
[249]DSA, Art 55.
[250]DSA, Art 50.
[251]Per Art 86, such bodies would encompass qualified entities under Directive (EU) 2020/1828 of the European Parliament and of the Council of 25 November 2020 on representative actions for the protection of the collective interests of consumers and repealing Directive 2009/22/EC [2020] OJ L409/1.

DSC in which the service recipient is established,[252] DSCs are also empowered with extensive investigatory and enforcement competences, which should be exercised in a 'effective, dissuasive and proportionate' manner.[253]

In terms of investigatory powers, DSCs can require service providers and traders to furnish information in relation to purported violations of the DSA, make applications for court orders allowing inspections of premises, or the seizure of information relating to purported violations, and ask employees and representatives of service providers to furnish information on purported violations.[254] In terms of enforcement, DSCs can accept and render binding (and enforceable) commitments of service providers in relation to compliance, render orders or impose remedies to stop violations of the DSA (or request a judicial authority or court to make such an order), impose fines or periodic penalties for failure to comply in line with Article 52 (or request a judicial authority or court to do so), or adopt interim measures.[255] Where the violation has not been remedied or continues, DSCs can require the management board of the service provider, firstly to elaborate a plan to stop the violation, monitor and report on actions taken; generally following consultation with interested parties, if the violation continues, the DSC can request a judicial authority to temporary restrict access to recipients of the service concerned, or where this is not technically feasible, to its online interface (a limited period which can be extended if necessary).[256]

Article 52 establishes that penalties imposed for non-compliance with the DSA must be effective, proportionate and dissuasive; failure to comply with obligations must not exceed 6M of annual worldwide turnover of the service provider, while period penalties must not exceed 5% of their average daily worldwide turnover.[257] Service recipients may also seek compensation from service providers for damage or loss incurred due to a violation of their DSA obligations.[258]

Violations of the OSA and DSA leading to cross-border harms

Cooperation between OFCOM and foreign regulators

The OSA somewhat recognises the likely cross-border nature of violations of online safety laws. Section 114 makes provision for cooperation and the

[252]DSA, Art 53. Member States in which services providers are established (or in which their legal representative is established – in the absence of such a connecting factor, the Commission is responsible per DSA, Art 56(7)) are generally understood to be competent in relation to supervision and enforcement, per Art 56 (save certain competences attributed to the Commission).
[253]DSA, Arts 51(5) and (6). Procedures and procedural rules fall within the procedural autonomy of the Member States, in line with applicable Union law, including the Charter of Fundamental Rights.
[254]DSA, Art 51(1).
[255]DSA, Art 52(2).
[256]DSA, Art 52(3).
[257]DSA, Art 52.
[258]DSA, Art 54.

sharing of information between OFCOM and foreign regulators. Nevertheless, the OSA treats non-compliance rather as a domestic issue. That is to say, it refers to the notion of cooperation or sharing of information to the ends of 'facilitating the exercise by the overseas regulator of that regulator's online regulatory functions', ie their online safety functions, and 'criminal investigations' relating to their regulatory functions.[259] It is worthwhile noting that there is an acknowledgment, both on the part of the OFCOM and the European Commission, that cross-border violations of online safety rules, are likely to arise.

As regards cooperation between UK and EU regulatory bodies, the Commission and OFCOM entered into an administrative agreement in May 2024, with the aim of supporting compliance with and enforcement of the OSA and DSA.[260] Few details on the framework of cooperation envisaged under the agreement have yet been published; thus far, it is anticipated that the cooperation will be facilitated through 'technical expert dialogues, joint training of technical staff, sharing of best practices, joint studies and coordinated research projects' to the ends of satisfying the objectives set out in each piece of legislation.

Cooperation between DSCs and the Commission

The powers of the Commission are defined contra the exclusive powers of Member States, per Article 56, though it is necessary that the Commission and Member State will cooperate. It is worth noting that the Commission has exclusive competence in relation to the supervision and enforcement of obligations applicable to VLOPs and VLOSEs under Section 5 of Chapter III.[261] It is anticipated that cooperation and mutual assistance will underpin enforcement; for example, DSCs shall inform other DSCs and the Commission when it exercises its competences.[262]

In respect of potential violations that have cross-border impacts, a DSC in one Member State can request the DSC of establishment to utilise their powers to ensure compliance with the DSA. If at least three DSCs consider there to be such a violations (ie in at least three Member States), the EBDS can request the DSC of establishment to utilise their powers to ensure compliance. Such requests must be reasoned, provide the contact details of the service provider, an outline of facts and the rationale underpinning the

[259]OSA, s 114(1). There are certain limitations to the sharing of such information; for example, it does not apply to intelligence service information (per section 116) or information that would breach data protection law, or that which is 'prohibited by Parts 1 to 7 or Chapter 1 of Part 9 of the Investigatory Powers Act 2016', per OSA, s 114(5).

[260]European Commission, 'Daily News, 13 May 2024', 3 <https://ec.europa.eu/commission/presscorner/api/files/document/print/en/mex_24_2568/mex_24_2568_en.pdf> accessed 01 June 2024.

[261]While Member States can exercise competences over VLOPs and VLOSEs established in their jurisdiction other than Section 5, Chapter III, where the Commission has not taken action – DSA, Art 56(4).

[262]DSA, Arts 56(7) and 57, and 49(2).

suspected violation alongside any other relevant information or suggestions as to measures that can be taken, and can be made unless the Commission has already initiated an investigation.[263] The DSC of establishment can request further information, if necessary, or can initiate a joint investigation under Article 60[264] and should provide a response to the DSC making the request, encompassing an assessment, evaluation of the purported violation and any measures taken.[265] If no response is given to the initial request, or the EBDS disagrees with the approach taken by the DSC of establishment, the EBDS can refer the matter to the Commission, which can undertake a further assessment; if the Commission considers the approach taken to be 'insufficient to ensure effective enforcement or [is] otherwise incompatible' with the DSA, the Commission can revert back to the DSC of establishment and request it reviews its initial assessment, in relation to which it must – taking into account the views of the Commission – engage its investigatory and enforcement powers and inform the Commission of its renewed approach.[266]

The scope for a joint investigation under Article 60 reinforces the significance of coordination and cooperation between DSCs to the enforcement of the DSA. A joint investigation can be initiated by the DSC of establishment where there is a suspicion of non-compliance affecting service recipients across multiple Member States, or on the EBDS's recommendation, following a request by at least three DSCs. The joint investigation may involve 'interested' DSCs; the DSC of establishment is characterised as the 'lead DSC'. It firstly reports, utilising its investigatory powers outlined above, on the purported violation and potential enforcement measures, taking into account the views of the other DSCs, to all DSCs, the Commission and the EBDS.[267] If the lead DSC fails to undertake the investigation, does not communicate its initial report, or if the EBDS disagrees with its position, the EBDS can refer the case to the Commission.[268]

While the Commission evidently has a role to play in relation to ensuring that all service providers comply with the DSA, as outlined above, arguably its key role pertains to the supervision and enforcement of the DSA in respect of VLOPs and VLOSEs as set out at Article 64 *et seq*. Again, underpinning these provisions is the notion of cooperation between the Commission, DSCs and the EBDS. They should work together to develop expertise and to identify and assess systemic risks arising from VLOPs and VLOSEs. The Commission can initiate its own investigation into purported violations

[263]DSA, Arts 58(1) to (3).
[264]DSA, Art 58(4).
[265]DSA, Art 58(5).
[266]DSA, Art 59(3).
[267]DSA, Arts 60(1) and (2).
[268]DSA, Art 60(3).

by VLOPs and VLOSEs on the basis of Article 66, while – in relation to purported Chapter III, Section 5 violations[269] – DSCs can request that the Commission investigates.[270]

Article 66 investigations also engage DSCs and the EBDS; having been informed of the investigation, the DSCs and EBDS should transmit relevant information to the Commission 'without undue delay'.[271] The requirement to investigate then shifts to the Commission; however, the Commission can request the support of the concerned DSCs or DSC of establishment, which must cooperate in a timely way and can utilise their Article 51 investigatory powers, while the Commission must inform those DSCs and the EBDS as to how it exercises its powers and its initial findings. In its decision, the Commission should take into account the views of the EBDS.[272]

The Commission's investigatory powers include powers to obtain information from the service providers, traders or auditors as well as DSCs; the request for information should be reasoned, indicate what information is required and its legal basis, the period in which it should be provided and the scope for the imposition of fines for 'incorrect, incomplete or misleading information'.[273] The Commission can also undertake interviews and request statements from relevant persons,[274] conduct inspections[275] and issue interim measures in cases of urgency and where there is a risk of serious damage.[276] Like DSCs, the Commission can accept and render binding commitments of VLOPs and VLOSEs leading to legally-binding decisions but retains the power to reopen its investigation if these commitments are not fulfilled, or reject such commitments if the Commission considers they will not lead to compliance.[277] Thus, the Commission can engage in monitoring of the service providers, also by appointing experts and auditors, to ensure any measures are implemented and ensure compliance.[278] The Commission's decisions on non-compliance must be preceded by a communication to the relevant service provider of its initial findings, rationale for decision, possible enforcement measures, and measures the service provider should take to effectively ensure compliance.[279] Decisions adopted will encompass an order requiring the service provider to take those measures, and indicate the relevant time period within which action should be taken;

[269]In respect of which the Commission has exclusive competence.
[270]DSA, Art 65.
[271]DSA, Arts 66(1) and (2) under the information sharing system set out in Art 85.
[272]DSA, Art 66(3) and (4).
[273]DSA, Art 67.
[274]DSA, Art 68.
[275]DSA, Art 69.
[276]DSA, Art 70.
[277]DSA, Art 71.
[278]DSA, Art 72.
[279]Moreover, the service provider and other persons have a right to be heard, and information on the Commission's file per DSA, Art 79.

the service provider should notify the Commission of the measure it has taken.[280] The Commission can also impose fines in this non-compliance decision on the service provider ('not exceeding 6% of its total worldwide annual turnover')[281] and can issue further fines for, for example, failing to comply with the Commission's investigatory powers.[282]

If the Commission adopts an Article 73 non-compliance decision, the service provider shall be subject to the 'enhanced supervision' provisions in Article 75, imposed by the Commission (taking account of the EBDS's opinion). This requires the service provider identify within a reasonable time, a plan and measures for compliance, to be communicated to the DSCs, Commission and EBDS. The EBDS should provide an opinion, and the Commission should decide if the measures ensure effective compliance, and thereafter, monitor its implementation, maintaining communicating with the EBDS and DSCs.[283] The Commission has the power to impose periodic penalties in line with Article 76 (of a maximum of 5% of the service provider's average daily income or worldwide annual turnover); these penalties may be for failing to comply with the Commission's investigatory powers or interim orders, legally-binding decisions adopted by the Commission or the action plan to correct non-compliance.[284] Commission decisions should be published,[285] and are subject to judicial review by the CJEU.[286] If compliance does not arise from the exercise of investigatory measures by DSCs or the Commission, the Commission can request the DSC of establishment to initiate Article 51(3) proceedings or initiate proceedings by itself, that is by seeking before a judicial authority the temporary restriction of access of recipients to the concerned service. While national courts should not take decisions that are counter to decisions already taken by the Commission, these decisions may still be subject to the preliminary reference procedure under Article 267 TFEU.[287]

Conclusions

This paper has endeavoured to illustrate that the emergence and development of digital platforms, and the sharing of content and services therein, raises a breadth of challenges in relation to the moderation of that content, whether it is illegal or lawful but harmful. In particular, it looks to unpack the nature and interests of the private service providers in

[280]DSA, Art 73.
[281]DSA, Art 74(1).
[282]DSA, Art 74(2).
[283]DSA, Art 75.
[284]DSA, Art 76. Arts 74 and 76 penalties, and their enforcement, are time limited per Arts 77 and 78.
[285]DSA, Art 80.
[286]DSA, Art 81.
[287]DSA, Art 82.

content moderation, and the duties and obligations established in the OSA and DSA in an attempt to regulate the systems and processes of online service providers. It is evident that the regulatory and governance frameworks envisaged under each generate uncertainties in relation to how these obligations can be enforced to, on the one hand, make internet safer, by mitigating risks of harm arising from the accessibility and dissemination of illegal and harmful content, while on the other, protecting users' fundamental rights and freedoms. While both pieces of legislation seek to satisfy similar aims, they adopt markedly different approaches to those ends, both in relation to the concept of harm engaged, and the regulatory framework adopted, but also as regards the procedural regimes established. Both the OSA and DSA encompass key procedural innovations, including internal complaints procedures and out-of-court dispute mechanisms, with the DSA providing a more comprehensive set of procedural rules; in both certain ambiguities and gaps remain (in relation to content moderation under the OSA and the vague concept of illegal content under the DSA), which may undermine the concrete impact envisaged. While it has been anticipated by lawmakers that both pieces of legislation will facilitate greater legal certainty and predictability for both service providers and users, their operation in practice remains to be seen.

Acknowledgements

I would like to extend my thanks to the anonymous reviewers for their helpful comments on the paper as well as to Hedvig Schmidt and Uta Kohl for constructive discussions on the topic.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Stephanie Law, Associate Professor of Law, University of Southampton, writes on EU law and fundamental rights, and has particular interests in consumer protection, data protection and procedural law.