A Medley of Public and Private Power in DSA Content Moderation for Harmful but Legal Content: An Account of Transparency, Accountability and Redress Challenges

by **Andrea Palumbo** *

Abstract: This paper analyses the challenges associated with the co-regulatory arrangement of Articles 34 and 35 of the Digital Services Act (DSA) for the mitigation of the risks posed by harmful but legal content. Filling a gap in the existing literature, this paper focuses on the implications of the public-private cogeneration of content moderation policies for harmful but legal content resulting from the implementation of the DSA. This paper highlights three challenges deriving from the "hybrid" public-private governance of harmful speech in the context of the DSA. First, the potential lack of transparency on public influence over private content moderation policies. Second, the risks of unaccountability of public bodies who are able to steer private content moderation policies to pursue public policy objectives. Third, the lack of effective redress available to users against interferences with their legal speech. Building on these considerations, this paper puts forward two main arguments. First, the governance model for online speech of the DSA poses challenges that are not addressed by the current legislative framework. Second, the public-private dichotomy of EU fundamental rights law is not fit for purpose in the face of hybrid regulatory models. Based on the identified challenges, the paper discusses the main shortcomings of the legislative framework, with the aim to define a path for future research.

Keywords: Digital Services Act, Content Moderation, Systemic Risks, Harmful Content, Freedom of Expression

Recommended citation: Andrea Palumbo, A Medley of Public and Private Power in DSA Content Moderation for Harmful but Legal Content: An Account of Transparency, Accountability and Redress Challenges, 15 (2024) JIPITEC 246 para 1

A. INTRODUCTION

1 The Digital Services Act ("DSA")[1] started to apply in its entirety on 17 February 2024. The DSA is widely regarded as a paradigm shift for the governance of online content, whose consequences for the dissemination of online content are yet to be fully understood. As the E-Commerce Directive, the DSA lays down rules for providers of mere conduit, caching and hosting services, which are grouped under the umbrella category of ''intermediary services".

2 By introducing new due diligence obligations for the providers of these services, the DSA moves from the model of *ex post* intermediary liability into the realm of both *ex ante* and *ex post* regulation.[2] Especially for providers of online platforms,[3] this paradigm

* Andrea Palumbo is Research Associate, Centre for IT & IP Law (CiTiP), KU Leuven.

1 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 (Digital Services Act) [2022] OJ L277/1.

2 Miriam C. Buiten, 'The Digital Services Act: From Intermediary Liability to Platform Regulation' (2021) JIPITEC 361.

3 According to Article 3(i) of the DSA, an online platform is a ''*hosting service that, at the request of a recipient of the service, stores and disseminates information to the public, unless that activity is a minor and purely ancillary feature of another service*

shift sees providers of intermediary services as more active, accountable and responsible actors in supervising the operation of their services and mitigating the risks they pose to fundamental rights, consumer protection and other societal and public interests.

3 In addition to entrusting new responsibilities to online intermediaries, the DSA imposes procedural obligations.[4]

4 The role given to procedural safeguards and fundamental rights in the DSA can be seen as a response that incorporates the dictates of digital constitutionalism.[5] Digital constitutionalism scholars have discussed the need to have in place solutions based on the rule of law to the challenges posed by the private ordering of online intermediaries,[6] and of online platforms.

5 The shift to platform regulation introduced by the DSA is clearest in relation to two categories of providers that, due to the heightened societal risks that may be caused by their services, are also subject to the highest standards of due diligence obligations. These are the providers of very large online platforms (''VLOPs'') and of very large online search engines (''VLOSEs'').

6 The due diligence obligations that Articles 34 and 35 of the DSA impose on VLOPs and VLOSEs are at the core of this paper. These Articles introduce risk assessment and mitigation obligations as well as a new mechanism of supervised regulation of online content. This new mechanism presents new questions and challenges that are investigated in the remainder of this paper.

7 Articles 34 and 35 DSA set up a new systemic risk mitigation system, in some respects translating a regulatory setting already tested in the field of the prudential supervision of financial institutions. While risk-based approaches are not new to EU digital regulation, systemic risk assessment and mitigation applied in relation to societal risks caused by technological design was an absolute novelty in EU legislation at the time of adoption of the DSA.

8 The focus of the analysis put forward in this article is on the content moderation policies enacted, under the supervision of the European Commission, by VLOPs and VLOSEs to mitigate the risks posed by legal content. By looking at the structure of the regulatory set-up in Articles 34 and 35 DSA, this paper argues that there is a phenomenon of cogeneration of content moderation policies resulting from the interaction between public and private actors, mainly because of the regulatory dialogue between the European Commission, on the one hand, and VLOPs and VLOSEs, on the other hand.

9 Based on this observation, this article describes the challenges that the public-private hybrid nature of decision-making processes for the moderation of legal content poses for transparency, accountability and effective redress for users whose freedom of expression may be limited. For the purposes of this article, the relevant legal framework against which these challenges are evaluated is EU law, with reference where appropriate to the European Convention on Human Rights (hereinafter, the ''ECHR'').

10 Existing literature has discussed the issues that arise in relation to public-private hybrid governance models for online content moderation. Most of the existing

or a minor functionality of the principal service and, for objective and technical reasons, cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means to circumvent the applicability of this Regulation".

4 These relate to content moderation decisions, and explicitly require providers to have regard for the fundamental rights of the recipients of their intermediary services. See Articles 17, 20 and 21 of the DSA.

5 See, among others: Nicolas Suzor, 'Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms' (2018) Social Media and Society 1; Edoardo Celeste, 'Digital Constitutionalism: A New Systematic Theorisation' (2019) International Review of Law, Computers & Technology 76; Giovanni de Gregorio, *Digital Constitutionalism in Europe: reframing rights & powers in the algorithmic society* (Cambridge University Press 2022); Oreste Pollicino, 'The quadrangular shape of the geometry of digital power(s) and the move towards a procedural digital constitutionalism' (2023) European Law Journal.

6 Joao Pedro Quintais, Naomi Appelman, Ronan Fahy, 'Using Terms and Conditions to apply Fundamental Rights to Content Moderation' (2023) German Law Journal 881, p. 907.

relevant literature precedes the DSA legislative proposal,[7] and discusses the issue in relation to the EU or national legal frameworks applicable, and the informal practices taking place, at the time. One contribution discusses the entanglements between public and private censorship across different legal frameworks, including the DSA.[8] These contributions generally discuss the phenomenon of public-private online content moderation. However, they do not address in detail how this phenomenon has been 'institutionalised' in the Digital Services Act, and which are the repercussions for the protection of users' freedom of expression when harmful but legal content is moderated under Articles 34 and 35 of the DSA. Therefore, this article intends to build on existing literature and further develop it by providing two novel contributions: i) an analysis of the public-private cogeneration of content moderation policies for harmful but legal content under the systemic risk assessment and mitigation regime of the DSA, and ii) an assessment of whether this regime poses risks for the effective protection of users' freedom of expression. The assessment in ii) is based on the evaluative criteria of transparency and accountability, defined in light of the legality principle of the Charter[9] for fundamental rights' limitations, and effective redress.

7 Michael D. Birnhack, Niva Elkin-Koren, 'The Invisible Handshake: The Reemergence of the State in the Digital Environment' (2003) Virginia Journal of Law and Technology 8(6); Derek E. Bambauer, 'Against Jawboning' (2015) Minnesota Law Review 182; Christopher T. Marsden, *Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace* (Cambridge University Press, 2011); Chris Marsden, Trisha Meyer, Ian Brown, 'Platform values and democratic elections: How can the law regulate digital disinformation?' (2020) Computer Law & Security Review 36 105373; Daphne Keller, 'Who Do You Sue? State and Platform Hybrid Power Over Online Speech' (2019) Hoover Institution, Aegis Series Paper No. 1902; Paddy Leerssen 'Cut Out By The Middle Man: The Free Speech Implications Of Social Network Blocking and Banning In The EU' (2015) JIPITEC 6(2) 99.

8 Rachel Griffin, 'The Politics of Algorithmic Censorship: Automated Moderation and its Regulation', in James Garratt, Music and the Politics of Censorship: From the Fascist Era to the Digital Age, Brepols.

9 Article 52 Charter of Fundamental Rights of the European Union [2012] OJ C 364/01.

11 The novelty of this article lies both in the specific focus on the DSA and in the evaluative framework adopted. It aims to answer the following research question: does the systemic risk assessment and mitigation regime of the DSA pose risks to the effective protection of users' freedom of expression about their harmful but legal content, in light of the evaluative criteria of transparency, accountability and availability of redress channels? The analysis is developed in different stages. Section 2 outlines the main features of the regime for systemic risk assessment and mitigation laid down in the DSA. This constitutes the starting point of the discussion, which is developed with a description in Section 3 of the regulatory dialogue and interaction between the Commission and supervised intermediaries that this regime entails. Section 4 discusses whether this dialogue would include some form of public interference with the freedom of expression of online users. Section 5 sets out the key passages of this contribution, discussing the challenges posed by the public-private cogeneration of content moderation policies for online harmful content in relation to transparency and accountability of public action, and to the redress mechanisms against interferences with harmful but legal content. Finally, Section 6 reflects upon the challenges of the hybrid speech governance model of the DSA, discussing potential solutions.

B. Mitigation of Systemetic Risks in the DSA: Articles 34 and 35

I. The Regime for Systemic Risk Assessment and Mitigation in the DSA

12 Articles 34 and 35 DSA feature a scheme for systemic risk assessment and mitigation which, on the one hand, give obliged entities wide discretion in assessing risks and designing measures to address them and, on the other hand, provides for mechanisms of continuous supervision and evaluation over the conduct of such entities.

13 Article 34 DSA requires VLOPs and VLOSEs to identify and assess the systemic risks arising in the Union

from the design, functioning and use of their services and related systems, carrying out risk assessments at least once a year.[10] Based on the systemic risks thus identified, VLOPs and VLOSEs shall put in place risk mitigation measures pursuant to Article 35 DSA.

14 The DSA provides for an articulate system of supervision over the conduct of VLOPs and VLOSEs. The European Commission is entrusted with the role of supervising the assessment and mitigation of systemic risks by VLOPs and VLOSEs, with pervasive supervisory and enforcement powers

15 The model embraced in the DSA combines the conferral of significant discretion to regulated entities with continuous supervision from a public body. It has been referred to as co-regulation[11] and meta-regulation.[12] The DSA is not the first piece of legislation to put in place a co-regulatory model, but it presents innovative features in relation to content moderation. Recital 104 of the DSA explicitly states that self and co-regulatory agreements, such as codes of practice, should be pursued to design risk mitigation measures, and the explanatory memorandum accompanying the legislative proposal of the DSA mentions the creation of a co-regulatory backstop in relation to the due diligence obligations for VLOPs and VLOSEs.[13]

16 It is beyond the scope of this paper to attribute a specific terminology to the regulatory approach of the DSA for VLOPs and VLOSEs. It suffices to note that the mixed public-private participation in the definition of rules governing the circulation of online content has been highlighted under different conceptualisations.

17 The *fil rouge* of such conceptualisations lies in the fact that, under the DSA, VLOPs and VLOSEs are under the responsibility to put in place self-regulatory measures to manage risks, whether in the form of implementing codes of conduct or other risk mitigation measures, while being under the supervision of another regulator that has the enforcement powers necessary to guide their conduct where appropriate.

II. Providers of Very Large Online Platforms and Search Engines as Risk Regulators for Online Content

18 In the regulatory scheme of the DSA for the mitigation of risks posed by the dissemination of illegal and harmful content, VLOPs and VLOSEs are attributed the role of risk regulators.[14] They are subject primarily to the oversight and enforcement by the European Commission, but also by the national Digital Services Coordinators (''DSCs'') and the European Board for Digital Services to a more limited extent.

19 Under Articles 34 and 35 DSA, VLOPs and VLOSEs have significant discretion in both phases of the systemic risk mitigation process, i.e. risk identification and mitigation. In particular, in recognition of their better-placed position to understand systemic risks and how to mitigate them, VLOPs and VLOSEs are the

10 The systemic risks to be assessed are not exhaustively determined *ex ante* in the DSA. Article 34 only provides a list of systemic risks that shall be assessed in any case, without precluding the identification of additional risks where appropriate.

11 Commission, 'Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC' COM(2020) 825 final, p. 3; Joan Barata, Oliver Budzinski, Mark Cole, Alexandre de Streel, Michèle Ledger, Tarlach McGonagle, Katie Pentney, Eleonora Rosati, 'Unravelling the Digital Services Act package' (2021), IRIS Special, European Audiovisual Observatory, Strasbourg, pp. 53, 54, 129; David Morar, ''The Digital Services Act's lesson for U.S. policymakers: Co-regulatory mechanisms'' [2022] Commentary published on the Brookings website, <https://www.brookings.edu/articles/the-digital-services-acts-lesson-for-u-s-policymakers-co-regulatory-mechanisms/> accessed 20 March 2024.

12 Nicolo Zingales, ''The DSA as a paradigm shift for online intermediaries' due diligence: hail to meta-regulation'', in Joris van Hoboken, João Pedro Quintais, Naomi Appelman, Ronan Fahy, Ilaria Buri, Marlene Straub (eds), ''Putting the Digital Services Act Into Practice: Enforcement, Access to Justice, and Global Implications'' (2023) Amsterdam Law School Research Paper No. 13, 2023, Institute for Information Law Research Paper No. 03, 2023.

13 Commission, 'Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC' COM(2020) 825 final, p. 3.

14 Nicolo Zingales (n 12) p. 216.

entities who bear primary responsibility for systemic risk identification and mitigation.

20 The risk-based approach provides a first legally binding tool to address the risks and harms of lawful but harmful content. VLOPs and VLOSEs must look both inward and outward, to identify and mitigate systemic risks caused not only by illegal content but also by other content.

21 In their role as risk regulators, VLOPs and VLOSEs may be called to take risk mitigation measures that have important consequences for freedom of expression. Article 35 of the DSA does not require, *per se*, to prohibit or restrict the dissemination of a specific category of content uploaded by recipients of the service, as the scope of its obligation is only the mitigation of any systemic risk identified pursuant to Article 34 of the DSA.

22 However, taking into account the rationale and the nature of the obligation in Article 35, its implementation can be expected to lead to content moderation policies that restrict speech protected by the right to freedom of expression. Besides the category of illegal content, Article 35 is aimed at addressing the risks posed by ''harmful'' content.

23 Harmful content is not defined in the DSA and is not a legal concept of EU law. It is a term widely used in policy discussions around platform regulation to indicate content that can create harm to a series of public and private protected interests, such as public security, public health, civic discourse and the physical and mental well-being of natural persons. A prominent example of content that may be legal but harmful is disinformation.

24 While the EU legislator purposefully avoided to define harmful content in the DSA, the connection between the systemic risk mitigation framework of the latter and the objective to regulate the dissemination of harmful content is evident.[15] This is clear from the recitals of the DSA that make reference to disinformation as information that may generate systemic risks[16] and, more generally, to content that may cause harm.[17]

25 In relation to harmful content, the DSA aims to address, *inter alia*, amplification-based harm that is caused when content is disseminated in a way that materialises a given level of systemic risks. The materialisation of systemic risks is the condition that triggers the application of Article 35 and, consequently, may justify the imposition of restrictions on harmful content.

26 The risk-based approach adopted by the EU legislator in the DSA, and the newly assumed role of risk regulators of certain providers of intermediary services, are a novelty in EU law that have been subject to criticism, especially for the risks they engender to freedom of expression. Concerns have been voiced in multiple respects.

27 First, the wording used by DSA provisions in describing the scope of application of Articles 34 and 35 of the DSA has been described as too vague. There has been criticism in relation to key terms such as ''systemic risks'', ''reasonable'' and ''proportionate'',[18] with potential deficiencies in meeting the legality test for the restriction of fundamental rights, in particular as concerns the foreseeability of future restrictions for online users.[19] As users must be able to foresee how their speech might be restricted to mitigate systemic risks, the wording of DSA provisions must satisfy minimum standards of clarity as to when content is deemed unacceptable and how it may be restricted.

15 Commission, 'Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC' COM(2020) 825 final, p. 3.

16 See recitals 9, 83, 84 and 104 of the DSA.

17 See recitals 5, 63, 69, 79, 104, 137 and 140 of the DSA.

18 Article 19, ''ARTICLE 19 recommendations for the Digital Services Act Trilogue'' (Article 19 website 2022) <EU: ARTICLE 19's recommendations for the Digital Services Act trilogue - ARTICLE 19> accessed 11 March 2024, pp. 2-3; Joan Barata, ''The Digital Services Act and its impact on the right to freedom of expression: special focus on risk mitigation obligations'' (2021) publication on *Plataforma en Defensa de la Libertad de Información* (PDLI), pp. 19 – 21; Joan Barata et al. (n 11) pp. 16-18.

19 Article 19, ''ARTICLE 19 recommendations for the Digital Services Act Trilogue'' (Article 19 website 2022) <EU: ARTICLE 19's recommendations for the Digital Services Act trilogue - ARTICLE 19> accessed 11 March 2024, pp. 2-3.

28 Second, the bestowal of regulatory functions, with wide discretion, to VLOPs and VLOSEs for the delicate task of moderating harmful but legal content has been criticised by the Committee on Civil Liberties, Justice and Home Affairs of the European Parliament as posing significant risks to freedom of expression. For this reason, during the legislative procedure for the adoption of the DSA the Committee proposed an amendment to delete the risk management framework.[20]

III. Article 35 of the DSA as Conducive to Restrictions on the Dissemination of Legal Content

29 The adoption of risk mitigation measures under Article 35 may lead to new types of interferences with freedom of expression targeting content that is legal and that is not restricted under any other legal basis than Article 35 itself. For illegal content, the boundaries of free speech have already been defined by national legislation or other legally binding acts, and the nature and scope of the restriction is clear as illegal content is plainly not acceptable.

30 Legal content whose dissemination generates systemic risks may instead be restricted under the aegis of Article 35 for the sole reason that it is conducive to such risks. Risk mitigation measures to be adopted under Article 35 are an open category, with the consequence that they may include anything deemed appropriate by VLOPs and VLOSEs.

31 For example, the prohibition of content is already foreseen as a possible measure to address disinformation in the Code of Practice on Disinformation.[21] Adherence to the Code may in turn constitute an adequate risk mitigation measure within the meaning of Article 35.[22] Another example is the demotion of content deemed harmful, which would fall under one of the measures listed in Article 35 regarding the adaptation of algorithmic systems, including recommender systems[23], and is equally foreseen as an action under the Code to tackle disinformation.[24]

32 Guidelines and reports published by the Commission to date in relation to systemic risks and the implementation of the DSA clearly show that the regulatory expectations towards Article 35 envisage the restriction of legal but harmful content. The Guidelines of the Commission on the mitigation of systemic risks for electoral processes recommend as mitigation measures, *inter alia*, disrupting the algorithmic amplification and spread of viral harmful content,[25] demonetisation[26] and the measures already foreseen in the Code on disinformation.[27] Furthermore, the report by the Commission on the application of the DSA risk management framework to Russian disinformation suggests in multiple points that restriction of legal speech may be warranted, especially as concerns the imposition of bans and demotions on Kremlin-aligned accounts.[28]

20 The Committee stated that the DSA ''*should address illegal content only and not "harmful content" as targeting legal content could put the freedom of expression at serious risk (i.e. annex to resolution 2020/2019(INL) as well as LIBE opinion PE650.375v02, par. 15), whereas the proposed Article 26 would go far beyond illegal content where mere vaguely described allegedly "negative effects" are concerned*''. See: Draft Opinion of the Committee on Civil Liberties, Justice and Home Affairs for the Committee on the Internal Market and Consumer Protection on the proposal for a regulation of the European Parliament and of the Council Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM(2020)0825 – C9-0418/2020 – 2020/0361(COD), Amendment 91, pp. 63-65.

21 The Strengthened Code of Practice on Disinformation [2022] measure 18.2, p. 20.

22 Article 35(1)(h) of the DSA, and preamble j) of the Code which states that signing up to all the commitments of the Code should be considered as a possible risk mitigation measure under the DSA.

23 Article 35(1)(d) of the DSA.

24 The Strengthened Code of Practice on Disinformation of 2022, measures 18.1 and 18.2, p. 20.

25 Commission, 'Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065', C/2024/2537, Section 3.2.1.

26 Ibid, p. 8.

27 Ibid, p. 5.

28 Commission, 'Digital Services Act: Application of the Risk Management Framework to Russian disinformation campaigns', <https://op.europa.eu/en/publication-

33 VLOPs and VLOSEs have already been moderating legal content before the DSA started to apply, in many cases on a voluntary basis.[29] Moreover, a search on the statements of reasons in the transparency database shows that some VLOPs demoted legal speech deemed harmful for civic discourse or elections relying on fully automated means.[30]

34 Under the DSA, VLOPs and VLOSEs may mitigate systemic risks by actively moderating legal content, and will be allowed to take a more proactive stance over monitoring content under Article 7 of the DSA. Compliance with Article 34 and 35 may require a rather systematic monitoring of the content being disseminated online. This could take place using online tools such as the demotion mechanisms in the ''Explore'' recommender system of Instagram,[31] where items may be filtered out or downranked by automated means based on integrity-related scores.[32]

C. Cogeneration of Content Moderation Policies in the DSA: Interaction Between the Commission and Supervised Intermediaries

35 As the entities in the best position to control the flow of online content, VLOPs and VLOSEs have been assigned important responsibilities for *ex ante* and *ex post* moderation of speech. The biggest change can be noted with regard to legal but harmful content, as for illegal content this was already the case before the DSA.[33]

36 When VLOPs and VLOSEs will continue to moderate legal content in compliance with Articles 34 and 35 of the DSA, the measures taken to this end cannot be deemed as merely private content moderation policies under their exclusive responsibility.[34] On the contrary, public authorities have the means and the duty to influence content moderation policies of obliged entities on a regular basis, where appropriate.

37 Thus, this article argues that the systemic risk mitigation framework of the DSA has set up a mechanism for the co-generation of the legal and technological rules that govern content moderation of online content, including legal content. On a legal level, cogeneration processes influence the terms and conditions on which VLOPs and VLOSEs rely as the contractual basis to restrict the content uploaded by service recipients. On a technological level, public actors can participate in shaping the algorithms governing the dissemination of online content by setting certain regulatory expectations on the implementation of Articles 34 and 35 of the DSA.

38 The public-private cogeneration of norms governing content moderation is a trend that has already emerged in recent years on an international level.[35]

detail/-/publication/c1d645d0-42f5-11ee-a8b8-01aa75ed71a1/language-en> accessed on 2 February 2024, pp. 45-46.

29 See the dashboard of the DSA transparency database, accessed at https://transparency.dsa.ec.europa.eu/dashboard on 2 February 2024.

30 See the results available at the following search on the DSA transparency database of the Commission: https://transparency.dsa.ec.europa.eu/statement?automated_decision%5B0%5D=AUTOMATED_DECISION_FULLY&category%5B0%5D=STATEMENT_CATEGORY_NEGATIVE_EFFECTS_ON_CIVIC_DISCOURSE_OR_ELECTIONS&platform_id%5B0%5D=36&platform_id%5B10%5D=31&platform_id%5B11%5D=34&platform_id%5B12%5D=30&platform_id%5B13%5D=22&platform_id%5B14%5D=27&platform_id%5B15%5D=29&platform_id%5B1%5D=28&platform_id%5B2%5D=23&platform_id%5B3%5D=37&platform_id%5B4%5D=32&platform_id%5B5%5D=24&platform_id%5B6%5D=25&platform_id%5B7%5D=26&platform_id%5B8%5D=33&platform_id%5B9%5D=35&page=19, accessed on 5 February 2024.

31 See the information available at the following link: https://engineering.fb.com/2023/08/09/ml-applications/scaling-instagram-explore-recommendations-system/, accessed on 6 February 2024.

32 For more information about how Meta uses AI to rank harmful content, see: https://ai.meta.com/blog/harmful-content-can-evolve-quickly-our-new-ai-system-adapts-to-tackle-it/, accessed on 6 February 2024.

33 Marco Bassini, 'Fundamental Rights and Private Enforcement in the Digital Age' (2019) European Law Journal 182.

34 Joan Barataet et al (n 11) p. 21.

35 Michael D. Birnhack, Niva Elkin-Koren (n 7); Derek E.

In the EU, it has been discussed more in detail in relation to the moderation of terrorist content,[36] under Regulation (EU) 2021/784.[37] Moreover, also the EU Code of Conduct on countering illegal hate speech online[38] introduced a collaboration between public and private entities for the moderation of illegal speech. This article aims to analyse this phenomenon with specific regard to the cogeneration of content moderation policies for content that is legal but can be restricted due to the systemic risks it poses. This scope of analysis thus warrants different considerations from content moderation policies that restrict content which is illegal under another national or EU law.

39 Private content moderation policies can be influenced by the European Commission and DSCs both *ex ante* and *ex post* under the DSA. The European Commission has exclusive competence to enforce the provisions on systemic risk assessment and mitigation for VLOPs and VLOSEs, while the competence is shared with the DSCs for enforcing compliance with the majority of the DSA provisions in relation to VLOPs and VLOSEs.

40 Articles 34 and 35 already define *ex ante* the basic rules governing content moderation for the purposes of mitigating systemic risks. They indicate how systemic risks must be assessed and lay down the high-level principles that should be respected when adopting mitigation measures, i.e. reasonableness, proportionality, effectiveness and consideration of the impact to fundamental rights.

41 However, due to the vagueness of the DSA provisions on systemic risk assessment and mitigation, the more concrete, and possibly consequential, *ex ante* guidance on how to implement them is likely to be provided by guidelines from, and dialogue with, the regulators.

42 First, the Commission, in cooperation with DSCs, can issue guidelines that provide more detail on how VLOPs and VLOSEs should mitigate systemic risks.[39] The first guidelines drafted under this legal basis have already been published, and they provide detailed guidance on how risks to electoral processes should be assessed and mitigated.[40]

43 Second, the Commission can invite VLOPs and VLOSEs to draw up codes of conduct on how to mitigate specific systemic risks.[41] The implementation of these codes can in turn qualify as a risk mitigation measure compliant with Article 35 of the DSA. The Commission has the means under the DSA to significantly influence the content of codes of

Bambauer (n 7); Christopher T. (n 7); Paddy Leerssen (n 7); Kylie Langvardt, 'Regulating Online Content Moderation' [2017] Georgetown Law Journal 1353; Daphne Keller (n 7); Julie E. Cohen, *Between Truth and Power* (Oxford University Press, 2019); Chris Marsden, Trisha Meyer, Ian Brown (n 7); Evelyn Douek, 'The Rise of Content Cartels' (publication on the website Knight First Amendment Institute at Columbia University, 2020) <The Rise of Content Cartels | Knight First Amendment Institute (knightcolumbia.org)> accessed on 7 February 2024.

36 Rocco Bellanova, Marieke de Goede, 'Co-Producing Security: Platform Content Moderation and European Security Integration' (2022) Journal of Common Market Studies 1316.

37 Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online (Text with EEA relevance) (2021) OJ L 172/79.

38 Code of conduct on countering illegal hate speech online (2019).

39 Article 35(3) reads as follows: ''*the Commission, in cooperation with the Digital Services Coordinators, may issue guidelines on the application of paragraph 1 in relation to specific risks, in particular to present best practices and recommend possible measures, having due regard to the possible consequences of the measures on fundamental rights enshrined in the Charter of all parties involved. When preparing those guidelines the Commission shall organise public consultations*''.

40 Commission, 'Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065', C/2024/2537, Section 3.2.1.

41 Article 45(2) of the DSA reads as follows: ''*where significant systemic risk within the meaning of Article 34(1) emerge and concern several very large online platforms or very large online search engines, the Commission may invite the providers of very large online platforms concerned or the providers of very large online search engines concerned, and other providers of very large online platforms, of very large online search engines, of online platforms and of other intermediary services, as appropriate, as well as relevant competent authorities, civil society organisations and other relevant stakeholders, to participate in the drawing up of codes of conduct, including by setting out commitments to take specific risk mitigation measures, as well as a regular reporting framework on any measures taken and their outcomes.*''

conduct, not only by encouraging their drawing up but also by monitoring how they are drafted[42] and how they are implemented.[43] In concrete terms, codes of conduct can become a set of rules implementing the DSA.

44 Third, the Commission may rely on Article 72(1) of the DSA to establish regular dialogue with VLOPs and VLOSEs on the mitigation of systemic risks.[44] In the context of this dialogue, the Commission can communicate regulatory expectations that may shape *ex ante* the measures to be adopted under Article 35.

45 Finally, the Commission and DSCs can influence the content and structure of terms and conditions in the context of their supervision over compliance by VLOPs and VLOSEs with Article 14 of the DSA, setting specific expectations on, for instance, clarity of the language and respect of fundamental rights.

46 *Ex post* intervention can take place thanks to the enforcement powers of public bodies that shape and constrain how VLOPs and VLOSEs deal with systemic risks.[45] Under Section 4 of Chapter IV of the DSA, the Commission has direct investigatory and enforcement powers over compliance of VLOPs and VLOSEs with the DSA. To this end, the Commission has a wide array of investigatory and enforcement powers at its disposal ranging from requests for information to inspections, interim measures, decisions of non-compliance and the imposition of fines.

47 Moreover, similarly to what is mentioned above regarding *ex ante* measures, Article 72(1) of the DSA can be relied on to have regular dialogue with supervised entities in order to correct and guide *ex post* their risk mitigation activities.

48 All of the enablers for *ex ante* and *ex post* intervention described above draw a picture of intricate relationships between EU public policy and private ordering of supervised entities. Public interference can affect the contractual freedom of private entities, as concerns the content of terms and conditions, as well as the freedom to structure the technological design of intermediary services. Contrary to other more subtle forms of government interference over content moderation induced through political or public opinion pressures,[46] the connection between private content moderation and public policy is made explicit in the DSA and stems from a legal requirement. While content moderation policies are ultimately determined by VLOPs and VLOSEs, lack of compliance with regulatory demands from the Commission or DSCs could lead to a decision of non-compliance and a fine. The enforcement activities of the Commission in the last year provide evidence of how subtly regulatory demands can be communicated to VLOPs and VLOSEs without the adoption of a formal decision. For example, in a letter dated 12 August 2024, former Commissioner Thierry Breton warned X owner Elon Musk about the dissemination of harmful content on X, with a specific mention of the upcoming live conversation with a US presidential candidate.[47] This shows that the *ex ante* and *ex post* avenues for intervention could also empower the Commission to put in place jawboning practices.

42 See Article 45(3) of the DSA.

43 See Article 45(4) of the DSA.

44 Article 72(1) of the DSA reads as follows: ''*for the purposes of carrying out the tasks assigned to it under this Section, the Commission may take the necessary actions to monitor the effective implementation and compliance with this Regulation by providers of the very large online platform and of the very large online search engines. The Commission may order them to provide access to, and explanations relating to, its databases and algorithms. Such actions may include, imposing an obligation on the provider of the very large online platform or of the very large online search engine to retain all documents deemed to be necessary to assess the implementation of and compliance with the obligations under this Regulation.*"

45 Joan Barata (n 11).

46 Niva Elkin-Koren, 'Government–Platform Synergy and its Perils' in Edoardo Celeste, Amelie Heldt, Clara Iglesias Keller (eds) *Constitutionalising Social Media* (Oxford: Hart Publishing, 2022), pp. 177–198.

47 The letter has been published by the X account of former Commissioner Thierry Breton on 12 August 2024. See: Thierry Breton on X: "With great audience comes greater responsibility #DSA As there is a risk of amplification of potentially harmful content in EU in connection with events with major audience around the world, I sent this letter to @elonmusk https://t.co/P1IgxdPLzn" / X.

49 The existence of legal obligations and a co-regulatory setting of dialogue with, and enforcement by, the Commission enables public policy considerations to affect private content moderation practice to a level that, before the DSA, was unprecedented in EU law. The extent to which public policy considerations seep into terms and conditions and technological design cannot be gauged from a reading of Article 35 of the DSA alone.

50 Obligations for systemic risk assessment and mitigation are defined in vague terms, and the concrete relationship between EU public policy and private ordering will become clearer only in the future when the Commission has consolidated its supervisory and enforcement practices. The Commission might issue detailed guidelines and encourage the drawing up of codes of practice to set out how harmful content should be moderated or have continuous dialogue with supervised entities indicating in a more informal manner which conducts are recommended.

51 After having described the setting that leads to the cogeneration of content moderation policies, a central aspect of this article is understanding the consequences of such setting for any restriction to harmful content. For illegal content, the nature of the restriction to free speech is clearly set out in the EU or national law that qualifies the content as illegal.

52 The nature and scope of the interference would result directly from the relevant legal provisions and their interpretation. The underlying constitutional calculus that justifies the illegality of certain speech has been made and rendered explicit by the authority that adopted such provisions, which in most cases would be the EU or national legislature. For instance, the terrorist content to be restricted under the Terrorist Content Regulation is clearly linked to the offences regulated by Directive 2017/541,[48] and many forms of hate speech are prohibited pursuant to the Council Framework Decision 2008/913/JHA.[49] In both of these examples the relevant constitutional calculus for the prohibition of certain categories of content has been made by the EU legislature.

53 For the restrictions to harmful content that may result from Article 35, the way the constitutional calculus takes place is more intricate and complex. Harmful and legal content is not prohibited as such by the DSA. For example, in the case of disinformation a single piece of false or misleading information is not per se the object of any legally-mandated restriction. However, if disseminated in a given manner and context that generates systemic risks of the type under the scope of Article 34 of the DSA, VLOPs and VLOSEs would be under an obligation to take mitigation measures that address this content, which may in turn involve measures that restrict its dissemination.

54 The manner and context of dissemination acquire central importance, as the same content may or may not have to be restricted depending on these factors. The example of amplification-based harm is instrumental in understanding the types of restrictions that may stem from Article 35 DSA. Certain content may become harmful only when amplified to a given extent that generates systemic risks, such as disinformation on health matters that may cause a public health crisis. The amplification-based harm caused by legal content is the element that may shift the constitutional calculus and induce the legislator to introduce an interference with freedom of expression.[50] This is the approach taken by the risk mitigation framework of the DSA, which only lays down the general principles governing risk assessment and mitigation without explicitly requiring restrictions to any category of content.

48 Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA [2017] OJ L 88/6.

49 Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law [2008] OJ L 328/55.

50 Daphne Keller, 'Amplification and its discontents: why regulating the reach of online content is hard' (Knight First Amendment Institute at Columbia University, section on essays and scholarship, 2021), <https://knightcolumbia.org/content/amplification-and-its-discontents> accessed on 3 March 2024.

55 The identification and mitigation of systemic risks becomes a crucial moment for the constitutional calculus under which an interference with freedom of expression is justified. Risk acts as a proxy for the balancing of conflicting constitutional interests:[51] the freedom of expression of users, public interests harmed by systemic risks, and the freedom to conduct a business of private entities. Therefore, to understand the role of cogeneration in shaping content moderation policies for legal content, it is essential to look at the decision-making process on the identification and mitigation of systemic risks. This process should be looked at taking into account the primary responsibility of VLOPs and VLOSEs in deciding how to assess and mitigate risks, but also the role of the Commission as meta-regulator to guide *ex ante*, and correct *ex post*, the actions of VLOPs and VLOSEs. This adds another venue of cogeneration besides content moderation policies, and it relates to the upstream task of systemic risk identification and mitigation.

56 Under the DSA, the Commission has a series of tools at its disposal to influence *ex ante* and *ex post* assessments of VLOPs and VLOSEs. In this context, Article 72 can be relied on to shape private assessments *ex ante*, whereas the investigatory and enforcement powers described above in relation to risk mitigation are equally used by the Commission to control *ex post* how systemic risks are identified and assessed.

57 Such intervention takes place *ex ante*, via the rules included in the DSA. It also takes place *ex post*, due to the capacity of the European Commission to shape and constrain the different ways platforms deal with systemic risks, which entail the dissemination of and access to far more types of content than merely illegal information.

58 Overall, a picture can be drawn where public and private actors cogenerate policies for the moderation of legal but harmful content. This cogeneration involves aspects of crucial constitutional relevance, notably the identification of the level of risk that may justify a restriction to freedom of expression and the determination of the nature and scope of such restriction.

59 As it determines the scope of the limitation on freedom of expression, the interplay between public influence and private ordering has important consequences for the freedom of expression of millions of recipients of VLOPs and VLOSEs' services. This interplay creates a new dimension of power and interference which may be difficult to categorise under traditional constitutional concepts. Thus, it creates new theoretical and practical challenges for the protection of the freedom of expression of users, as is discussed below.

D. Public Interferences with Freedom of Expression under the DSA: Article 35 and Public Supervision over Private Ordering

60 The rights recognised by the Charter of Fundamental Rights of the European Union are binding on the EU institutions and the Member States when they implement EU law, as provided for in its Article 51.[52] Whether the Charter can impose horizontal direct effects on individuals is an issue that has been settled in EU case-law in relation to certain rights,[53] but remains open as concerns other rights including freedom of expression.[54] Therefore, the most common and clear application of the Charter is as standard of review of public actions. For the Charter to be invoked against a public action, this action must qualify as a public interference with a fundamental right protected by the Charter.

51 Giovanni De Gregorio, Pietro Dunn, 'The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age' (2022) Common Market Law Review, pp. 17-18.

52 Article 51 of the Charter reads as follows: "*the provisions of this Charter are addressed to the institutions and bodies of the Union with due regard for the principle of subsidiarity and to the Member States only when they are implementing Union law. They shall therefore respect the rights, observe the principles and promote the application thereof in accordance with their respective powers.*"

53 ECJ Case C-414/16 *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung eV* [2018] EU:C:2018:257; ECJ Case C-569/16 *Stadt Wuppertal v Maria Elisabeth Bauer and Volker Willmeroth v Martina Broßonn* [2018] ECLI:EU:C:2018:871.

54 Paul Craig, Grainne de Búrca, *EU Law: Text, Cases, and Materials* (seventh edition, Oxford University Press, 2020), pp. 450-454.

61 The public influence on private content moderation policies raises the question of whether a risk mitigation measure taken under Article 35 of the DSA can qualify as a public interference under the Charter. To answer this question, it is necessary to look at the entire co-regulatory setting for systemic risk mitigation of the DSA, with the different layers of legislative norms, supervisory and enforcement actions and private conduct.

62 The case-law of the ECJ and of the European Court of Human Rights (ECtHR) offers a rather clear picture of what can constitute an interference with freedom of expression. The ECtHR has developed a broad interpretation of what could constitute an interference, including a large array of measures such as formalities, conditions, restrictions or penalties.[55] On the internet, restrictions on the means of dissemination are interferences with freedom of expression as much as restrictions on content per se,[56] especially when they target online content-sharing platforms that play a vital role for the dissemination of information.[57]

63 Therefore, restrictions that affect recommender systems and the algorithmic reach of online content may qualify as an interference with freedom of expression. According to the case-law of the ECtHR, an interference does not need to result from a legally binding act. It may also stem from other actions of a public authority which have the effect of restricting the enjoyment of freedom of expression, potentially also in the form of chilling effects.

64 The case-law of the ECJ sheds light on forms of public interference where the restriction to free speech stems indirectly from public action, even in the absence of a direct prohibition of a given content. The reasoning followed by the ECJ in Poland v EP and Council[58] shows that a provision can qualify as an interference with freedom of expression even if it does not directly prohibit content nor explicitly requires intermediaries to restrict content. It suffices that the provision generates a situation that would lead intermediaries to restrict content to comply with it. In particular, the ECJ held that, since Article 17 of Directive 2019/790[59] ("DSM Directive") created a situation where intermediaries would need to deploy content filtering tools that may block *ex ante* legal content, the limitation resulting from these tools would be attributable to the EU legislature.[60]

65 This attribution flows from the fact that the limitation is a direct consequence of the specific liability regime established in respect of online content-sharing service providers in Article 17(4) of the DSM Directive. As a consequence, the ECJ concluded that the specific liability regime of Article 17(4) of the DSM Directive entails a limitation on the exercise of the right to freedom of expression and information of users guaranteed in Article 11 of the Charter.[61]

66 The principles described above on the meaning of public interference provide convincing arguments to claim that there is a public interference with freedom of expression when the application of Article 35 leads to the moderation of legal content. At the legislative level, Articles 34 and 35 provide the basis and the metrics to put in place such restrictions. Since legal but harmful content is not restricted under other laws that declare its illegality, the only criteria that determine its restriction are those stemming from the risk management regime of the DSA, i.e. the

55 *Wille v. Liechtenstein* App no 28396/95 (ECtHR 28 October 1999), para. 43.

56 *Autronic AG v. Switzerland* App No 12726/87, (ECtHR 22 May 1990), para. 47; *Murphy v. Ireland* App No 44179/98 (ECtHR, 10 July 2003), para. 61; *Ahmet Yildirim v. Turkey* App No 3111/10 (ECtHR 18 December 2012), para. 50; *Pirate Bay: Neij and Sunde Kolisoppi v Sweden* App No 40397/12 (ECtHR 18 February 2013); *Pendov v. Bulgaria* App No 44229/11, (ECtHR 12 October 2020), para. 53.

57 *Cengiz and Others v. Turkey* Applications nos. 48226/10 and 14027/11 (ECtHR 1 December 2015), para. 52; *Vladimir Kharitonov v. Russia* App No 10795/14 (ECtHR 23 June 2020), para. 33 and the case-law cited therein.

58 ECJ Case C-401/19 *Republic of Poland v. European Parliament and Council of the European Union* [2022] EU:C:2022:297.

59 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 L 130/92.

60 ECJ Case C-401/19 *Republic of Poland v. European Parliament and Council of the European Union* [2022] EU:C:2022:297, para. 56.

61 ibid para. 58.

capacity to create systemic risks. And while Article 35 does not directly restrict content, it establishes duties of care that oblige intermediaries to impose limitations on legal speech in certain circumstances. The limitation is indirect but directly attributable to the EU legislature, following the same logic of the ECJ in Poland v EP and Council.

67 At the supervisory and enforcement level, the Commission has the means to influence private policies interfering with freedom of expression in various ways. The Commission can contribute to the definition of content moderation policies by providing *ex ante* guidance and by correcting private actions *ex post*. The high-level and vague wording of Articles 34 and 35 of the DSA provide the basis and metrics to restrict harmful content, but the Commission can shape the rules that govern the moderation of legal content in practice. The actual decisions on which risks are acceptable, and which should be mitigated and how, is a result of the intricate relationship between public supervision/enforcement and private ordering. As discussed above, VLOPs and VLOSEs as risk regulators carry out a balancing exercise of fundamental importance together with the Commission.

68 In light of the above, there are elements to argue for the existence of a mediated public interference with freedom of expression. This interference would result from Article 35 of the DSA and, possibly, from the supervisory and enforcement action of the Commission. If this is the case, the Charter would apply. The existence of a public interference can therefore not be excluded solely for the fact that private entities are ultimately responsible for moderating content that generates systemic risks.

E. Transparency, Accountability and Redress for Public Interferences with Freedom of Expression

69 This article argues that the mix of private and public participation in the definition of content moderation policies for online legal but harmful content raises specific concerns for the effective protection of the freedom of expression of users. This conclusion is based on an assessment that relies on two evaluative criteria: i) transparency and accountability in relation to public interferences with freedom of expression, and ii) judicial redress channels available against such interferences. These two concerns are described below.

I. Transparency and Accountability of Public Interferences with Freedom of Expression under the DSA

70 Transparency and accountability are widely recognised as foundational principles of good governance, essential to build trust in public actors.[62] They are also guiding principles of pervasive importance for the actions of the European Commission,[63] and recognised pillars of the rule of law in EU legislation[64] and policy[65] and in the Council of Europe.[66] Transparent and accountable actions by

62 Michael Johnston, 'Good governance: Rule of law, transparency, and accountability' (United Nations Public Administration Network, 2006), <Good governance: rule of law, transparency, and accountability | IIEP Unesco - Etico | Platform on ethics and corruption in education> accessed 6 March 2024; European Parliament, 'Transparency, integrity and accountability in the EU institutions' (briefing for the PETI Committee, 2019) <Transparency, integrity and accountability in the EU institutions (europa.eu)> accessed on 10 January 2024, p. 1; Council of Europe, '12 principles of good democratic governance', (online brochure, 2019), <https://rm.coe.int/brochure-12-principles-of-good-governance-and-current-tools-on-good-go/16808b1687> accessed on 10 January 2024; Janos Bertok, 'Public Sector Transparency and Accountability: Making it Happen' (OECD/OAS, OECD publishing Paris, 2002).

63 Commission, ''Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Better regulation for better results - An EU agenda'', COM(2015)215 final.

64 According to Article 2(a) of Regulation 2020/2092, the 'rule of law' includes ''*the principles of legality implying a transparent, accountable, democratic and pluralistic law-making process*''.

65 Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2020 Rule of Law Report, The rule of law situation in the European Union', COM(2020) 580 final.

66 European Commission for Democracy through Law (Venice Commission), 'Rule of law checklist' (publication of the

the EU public administration are also instrumental to enable citizens to effectively exercise their rights, by monitoring how they may be affected by public action.

71 The DSA provides for increased transparency about the activities of intermediary service providers and enhances their accountability vis-à-vis recipients of their services.[67] As concerns content moderation, the DSA represents a significant step forward towards transparency and accountability regarding the restrictions on content put in place by providers of intermediary services and the underlying reasons for such restrictions.[68]

72 The DSA provides for articulated transparency safeguards provided against private restrictions on content. However, there are no mechanisms to ensure that users can clearly distinguish between measures that are purely private and measures that providers are obliged to adopt to tackle systemic risks under the DSA. Ultimately, it seems that the DSA does not fare as well when it comes to providing for transparency and accountability over public intervention in the dissemination of information on online platforms and search engines. In particular, issues can be identified regarding the foreseeability of interferences with freedom of expression from the perspective of online users. Even though Article 35 does not create obligations for users, but only for platforms, it is the legal basis for interferences with freedom of expression that may be imposed on the harmful but legal content of users. Therefore, foreseeability should be looked at from the perspective of online users whose legal content may be restricted as a consequence of Article 35.[69] Foreseeability is an essential quality that any legal basis providing for an interference with fundamental rights should have,[70] in order to be compliant with the requirement under the Charter that such interference be 'provided for by law'.[71]

73 While the difficulty to distinguish private and public censorship in contemporary trends of content moderation has already been observed[72], the DSA creates new and more concerning challenges in relation to the moderation of legal but harmful content. Article 14(1) of the DSA does not require

Council of Europe, 2016).

67 The DSA has provisions on both *ex ante* and *ex post* transparency. The former is ensured by Article 14(1) of the DSA, according to which terms and conditions must inform users of any restrictions that may be imposed in relation to the use of an intermediary service, including information on how content moderation takes place for any content that is illegal or incompatible with terms and conditions. The same provision also lays down requirements over how this information should be communicated, i.e. that it be in clear, plain, intelligible, user-friendly and unambiguous language, publicly available in an easily accessible and machine-readable format. The latter is enabled by provisions on statements of reasons accompanying decisions that restrict online speech, by the obligation to publish reports with information on the content moderation activities that intermediaries engaged in, and by the obligation to make publicly available the specific mitigation measures put in place to address systemic risks.

68 Giancarlo Frosio, Christophe Geiger, 'Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime' [2023] European Law Journal 31; Joao Pedro Quintais, Naomi Appelman, Ronan Fahy, 'Using Terms and Conditions to apply Fundamental Rights to Content Moderation' [2023] German Law Journal 881.

69 This is, in essence, the perspective adopted by the ECJ in assessing whether the liability regime in Article 17 of Directive 2019/790 imposes a limitation on the exercise of the right to freedom of expression and information that is compliant with the Charter. See: ECJ Case C-401/19 *Republic of Poland v. European Parliament and Council of the European Union* [2022] EU:C:2022:297.

70 European Union Agency for Fundamental Rights, 'Applying the Charter of Fundamental Rights of the European Union in law and policymaking at national level' [2020], pp. 71-72; ECJ Case C-401/19 *Republic of Poland v. European Parliament and Council of the European Union* [2022] EU:C:2022:297, para. 67; ECJ joined Cases C-203/15 and C-698/15 *Tele2 Sverige AB and Secretary of State for the Home Department* [2016] EU:C:2016:970, para. 117; Opinion of Advocate General Cruz Villalón in ECJ Case C-70/10 *Scarlet Extended SA v SABAM* [2011] EU:C:2011:255, paras. 94-96. Opinion in ECJ Case C-401/19 *Poland v. Parliament and Council* [2022] ECLI:EU:C:2022:297, para 90.

71 According to Article 52(1) of the Charter, any interference with the enjoyment of the rights and freedoms recognised by the Charter must be 'provided for by law'. This requirement is also known as the legality principle, and is one of the conditions that must be met for an interference with a fundamental right to be compliant with the Charter.

72 Rachel Griffin, 'The Politics of Algorithmic Censorship: Automated Moderation and its Regulation', in James Garratt, *Music and the Politics of Censorship: From the Fascist Era to the Digital Age*, Brepols.

providers of intermediary services to inform users *ex ante* about which restrictions may apply specifically for the purposes of mitigating systemic risks under the DSA. Moreover, it does not require to indicate whether any specific measure applied has been required by the Commission through guidelines or informal supervisory requests. While this provision safeguards the foreseeability of the restrictions that may be imposed because of the contractual document governing the relationship with providers, i.e. terms and conditions, it does not ensure full transparency in relation to the restrictions that derive from an application of Article 35.

74 When reading terms and conditions, a user may not be able to distinguish between restrictions that result from the contractual freedom of providers or from demands of public policy. All interferences with content that is legal may be equally enforceable and justified as incompatible with terms and conditions. Therefore, any restriction with harmful content would fall under the same contractual ground and might, *prima facie*, appear as resulting from a decision solely of the service provider. Moreover, Article 17 on statements of reasons only obliges providers to indicate the contractual ground relied on for the restriction of legal content. It does not require an indication of whether the interference is a risk mitigation measure implemented in compliance with Article 35 of the DSA. While for illegal content the basis of the restriction would be clearly set out in law, and would have to be specifically indicated in statements of reasons, for legal content there is no transparency on the nature of the public interference behind private content moderation.

75 Not only is there no guarantee of transparency from VLOPs and VLOSEs on this point, but also the text of Article 35 is excessively vague to predict *ex ante* which content may be restricted for reasons of public policy due to the fact that it generates systemic risks,[73] as there is no clear criteria to determine when a risk becomes too risky.[74]

76 As discussed above, the public-private cogeneration of content moderation measures is the context where conflicting constitutional interests in relation to harmful but legal content are balanced. These reasons may not be made public, and the user may not have means to understand to which extent public policy considerations are behind what is and is not accepted on online fora.

77 The biggest promise for transparency may lie in Article 42(4)(b), according to which VLOPs and VLOSEs must make publicly available, at least once a year, the specific mitigation measures put in place pursuant to Article 35(1). However, the level of granularity in which information on mitigation measures is disclosed will be the key factor in effectively enabling users to be aware of which restrictions applied to them stem from such measures. Moreover, the requirement to publish information on mitigation measures once a year does not enable *ex ante* foreseeability of legally mandated censorship over legal content, but only provides *ex post* reporting. Overall, a mechanism where private entities formally take decisions that *de facto* have been required by law and influenced by public bodies does not provide for transparency over public interferences with freedom of expression. As a consequence, it may enable public bodies to evade accountability for their online speech policies. Online users, and society at large, would not be able to have detailed information about how systemic risks have been assessed, how the appropriate mitigation measures to be implemented have been determined, and which restrictions to harmful content provided for in the terms and conditions of VLOPs and VLOSEs are mitigation measures put in place according to Article 35(1).

73 Article 19, ''ARTICLE 19 recommendations for the Digital Services Act Trilogue'' (Article 19 website 2022) <EU: ARTICLE 19's recommendations for the Digital Services Act trilogue - ARTICLE 19> accessed 11 March 2024, pp. 2-3; Joan Barata (n 11) pp. 19–21; Joan Barata et al. (n 11) pp. 16-18.

74 Joan Barata, ''The Digital Services Act and its impact on the right to freedom of expression: special focus on risk mitigation obligations'' (2021) publication on *Plataforma en Defensa de la Libertad de Información* (PDLI), p. 20.

78 While not desirable for reasons of transparency and accountability, it is not clear to which extent this mechanism is also in violation of EU law provisions governing public action. While the delegation of public tasks to private entities can be criticised from the perspective of constitutional legitimacy,[75] a sharing of responsibilities between public and private entities is also widely regarded as essential in a complex world.[76] This is especially the case where private entities are in the best position to address certain societal issues such as harmful content.

79 From the perspective of the requirement that any interference with fundamental rights be ''provided for by law'', the legal basis providing for an interference with fundamental rights does not need to be adopted by a democratically legitimised body.[77] It suffices that it is an act of general application with the requisite quality of ''law''.[78] Moreover, it is not precluded that norms of private entities provide for the interference insofar as there is a delegation from public bodies to this end and there is appropriate public oversight over how the delegated powers are exercised.[79]

80 Nonetheless, it can be argued that the blurring line between public and private censorship for legal content is problematic for both legality *stricto sensu*, on the one hand, and for accountability on the other. The two aspects are interconnected, as they are both essential components of the rule of law.[80]

81 Lack of transparency over public action in content moderation does not enable users to fully understand the nature of the limitations to their fundamental rights and, consequently, it creates a gap in the accountability of public bodies. The lack of accountability and transparency may in turn increase the risks of an arbitrary exercise of power by public bodies, whose actions entailing a public interference with freedom of expression may not be recognisable and such. This may ultimately thwart the objective of the legality principle to function as a safeguard against arbitrariness of public action, which implies that public interferences are clearly recognisable as such and foreseeable in relation to their effects.

II. Redress channels against interferences with legal speech

82 The DSA is a real breakthrough when it comes to the protection of online users through procedural safeguards and redress mechanisms. It translates rule of law principles to the governance of online platforms, recognising the role of large providers of intermediary services as *de facto* rule setters with powers comparable to that of a state entity, as theorised by multiple scholars to date.[81] Examples

75 Rikke Frank JØrgensen (ed.), *Human Rights in the Age of Platforms* (The MIT Press, 2019).

76 Robert Baldwin, 'Better Regulation: The Search and the Struggle' in Robert Baldwin, Martin Cave, Martin Lodge (eds), *The Oxford Handbook of Regulation* (Oxford, Oxford University Press 2010), pp 259–278; Neil Gunningham, Darren Sinclair, 'Smart Regulation' in P Drahos (ed.), *Regulatory Theory: Foundations and Applications* (Canberra, ANU Press 2017); Christine Parker, 'Twenty Years of Responsive Regulation: An Appreciation and Appraisal' (2013), Regulation & Governance 2; Almada Marco, 'Regulation by Design and the Governance of Technological Futures' (2023) European Journal of Risk Regulation.

77 Robert Schutze, *European Union Law* (Oxford University Press, 2015), pp. 446-447; Koen Lenaerts, 'Exploring the Limits of the EU Charter of Fundamental Rights' (2021) European Constitutional Law Review, pp. 388-391;

78 Ibid.

79 *Barthold v Germany* App No 8734/79 (ECtHR 25 March 1985), para. 46; Hans-Bredow-Institut for Media Research, 'Study on Co-Regulation Measures in the Media Sector' (University of Hamburg, Final Report, Study for the European Commission, Directorate Information Society and Media, 2006) pp. 147-152.

80 Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2020 Rule of Law Report, The rule of law situation in the European Union', COM(2020) 580 final; European Commission for Democracy through Law (Venice Commission), 'Rule of law checklist' (publication of the Council of Europe, 2016).

81 Nicolas Suzor, 'The Role of the Rule of Law in Virtual Communities' (2010) Berkeley Technology Law Journal 1817; Niva Elkin-Koren, Maayan Perel, 'Guarding the Guardians: Content Moderation by Online Intermediaries and the Rule of Law' in Giancarlo Frosio (ed.), *Oxford Handbook of Online Intermediary Liability* (Oxford University Press, 2020); Stephan Kolo ßa, 'Facebook and the Rule of Law' (2020) Zeitschrift für ausländisches öffentlichesRecht und Völkerrecht 509;

of procedural safeguards and redress mechanisms include transparency requirements,[82] the obligation to state reasons,[83] the internal complaint-handling system[84] and out-of-court dispute settlement.[85] These safeguards aim at protecting users against arbitrary and non-transparent content moderation practices, and at holding providers of intermediary services accountable for their actions.

83 Nonetheless, the remedies available to users against content moderation actions of private entities are different from those that would be available if the restriction of content were the clear result of public action.

84 If VLOPs and VLOSEs are enabled by the terms and conditions governing their services to restrict the dissemination of legal but harmful online content, users may only be able to act against such restriction by claiming that it is in violation of the terms and conditions. For example, there may be cases where online content is restricted on the basis of it being incompatible with terms and conditions. Should users believe that the lawful content was wrongfully labelled as incompatible, they may submit a complaint to the provider, or alternatively act before an out-of-court dispute settlement body or a judge. However, any such claim of the user against VLOPs and VLOSEs may only be based on contractual grounds, and in particular on the terms and conditions that regulate their private relationship.

85 Besides any claim based on the contractual norms governing the relationship between users and providers, it is not clear whether users could demand that VLOPs and VLOSEs respect their fundamental right to freedom of expression. In particular, if they could require that any interference with legal speech be justified in accordance with the criteria laid down in Article 52(1) of the Charter. Admitting this type of legal action would be tantamount to recognising the horizontal direct applicability of the right to freedom of expression in private relationships.

86 In legal doctrine, the horizontal direct effect of a provision denotes its ability to find application in cases between private parties, with the consequence that a private party may rely directly on that provision in judicial proceedings against another private party.[86]

87 Giving direct effect to fundamental rights in contractual relationships means that they apply directly in such relationship in the same way they do in a state (or EU institution or body) – citizen relationship,[87] and therefore that fundamental rights act as a direct limitation to the freedom of contract of the parties. The ECJ has explicitly stated that Article 51(1) of the Charter should not be interpreted as systematically precluding the horizontal effect of fundamental rights,[88] thus leaving the question open for the future.

88 Despite the fact that the Charter is not addressed to private parties, the ECJ has recognised the

Nicolas Suzor, 'Digital Constitutionalism: Using the Rule of Law to Evaluate the Legitimacy of Governance by Platforms' (2018) Social Media and Society 1; Edoardo Celeste, 'Digital Constitutionalism: A New Systematic Theorisation' (2019) International Review of Law, Computers & Technology 76; Giovanni de Gregorio, *Digital Constitutionalism in Europe: reframing rights & powers in the algorithmic society* (Cambridge University Press 2022); Oreste Pollicino, 'The quadrangular shape of the geometry of digital power(s) and the move towards a procedural digital constitutionalism' (2023) European Law Journal.

82 See Article 14(1) of the DSA.

83 See Article 17 of the DSA.

84 See Article 20 of the DSA.

85 See Article 21 of the DSA.

86 Paul Craig, Grainne de Búrca, *EU Law: Text, Cases, and Materials* (seventh edition, Oxford University Press, 2020), pp. 225-232.

87 Chantal Mak, 'Fundamental rights in European Contract Law' (Dphil thesis, University of Amsterdam 2007), p. 49.

88 ECJ joined cases C-569/16 and C-570/16 *Stadt Wuppertal v Maria Elisabeth Bauer and Volker Willmeroth v Martina Broßonn* [2018] ECLI:EU:C:2018:871, para 87. In particular, the ECJ held in para 87 that: ''*although Article 51(1) of the Charter states that the provisions thereof are addressed to the institutions, bodies, offices, and agencies of the European Union ... and to the Member States only when they are implementing EU law, Article 51(1) does not, however, address the question of whether those individuals may, where appropriate, be directly required to comply with certain provisions of the Charter and cannot, accordingly, be interpreted as meaning that it would systematically preclude such a possibility*".

horizontal direct effect of certain fundamental rights in private relationships on account of their mandatory and unconditional nature,[89] namely the right to non-discrimination,[90] the right to effective judicial protection[91] and the right to paid annual leave.[92] Therefore, it appears that the recognition of direct horizontal applicability is specific to each fundamental right based on its nature.

89 The horizontal application of freedom of expression has for long been a controversial issue, and it has been discussed, among others, specifically in relation to the contractual relationship between platforms and users.[93] To date, however, there is no judgement of the ECtHR or of the ECJ in legal proceedings brought by a user against an online platform for restrictions on content. Thus, there is no recognition of the horizontal effect of freedom of expression in such a relationship.

90 In the absence of a clear judicial recognition of the horizontal effect of freedom of expression in contractual relationships in the EU,[94] it cannot be claimed with sufficient certainty that online users would be able, at the EU level, to act against any content moderation decision that VLOPs and VLOSEs implement to mitigate the systemic risks caused by harmful but legal content under Article 35 of the DSA. At the national level, there is no common trend among Member States to recognise the horizontal effect of freedom of expression in private relationships, despite the ruling by the German Federal Court of Justice that online intermediaries face constitutional obligations vis-à-vis their users.[95]

91 If an interference with freedom of expression resulted from a public action, however, users would be able to directly invoke their right under the Charter in a legal action against a public body. For example, a user may bring action against a legal act of the European Commission that imposes a restriction on their online speech. To this end, they may claim that there is an unjustified interference with their right to freedom of expression, e.g. because the interference is in violation of the principle of proportionality. In this case, users would act in a terrain with significant legal certainty, given that the fundamental rights of the Charter have been primarily applied in private-public relationships. There is ample case-law of the ECJ and the ECtHR clarifying under which conditions an interference by public bodies with freedom of expression may be considered lawful.

92 In relation to harmful but legal content, any restriction that can be attributed to a public body may be challenged by users on multiple grounds, including that the systemic risks caused by the content in question do not justify an interference with its dissemination.

93 When VLOPs and VLOSEs restrict the dissemination of harmful but legal content to mitigate systemic risks in compliance with Article 35 of the DSA, the source of the interference is private contractual law, and in particular terms and conditions. Even when the regulatory dialogue between the European Commission and VLOPs and VLOSEs leads to the coproduction of content moderation decision, as described above, the interference would still stem from a private measure. There is no public action that can be *prima facie* connected to it. There would be no act from the European Commission that specifically mandates the restriction of specific content, as the

89 Paul Craig, Grainne de Búrca (n 87) pp. 225-232.

90 ECJ Case C-414/16 *Vera Egenberger v Evangelisches Werk für Diakonie und Entwicklung eV* [2018] EU:C:2018:257, paras 76-82.

91 Ibid.

92 ECJ joined cases C-569/16 and C-570/16 *Stadt Wuppertal v Maria Elisabeth Bauer and Volker Willmeroth v Martina Broßonn* [2018] ECLI:EU:C:2018:871, para 85.

93 Matthias C. Kettemann, Anna Sophia Tiedeke 'Back up: Can users sue platforms to reinstate deleted content?' (2020) Internet Policy Review 1.

94 It is worth recalling that, as far as the EU legal order is concerned, the horizontal applicability of fundamental rights in the relationship between online platforms and users is an open question, as explicitly stated by Advocate General Saugmandsgaard Øe in his Opinion in ECJ Case C-401/19 *Poland v. Parliament and Council* (2022) ECLI:EU:C:2022:297, para 90.

95 German Federal Supreme Court, Judgements of 29 Jul 2021, III ZR 179/20 and ZR 192/20; see also Matthias C Kettemann, Torben Klausa, 'Regulating Online Speech: Ze German Way' (Lawfare, 20 September 2021) https://www.lawfareblog.com/regulating-online-speech-ze-german-way accessed 15 March 2024; Matthias C. Kettemann, Anna Sophia Tiedeke (n 93).

ultimate decision on how to mitigate systemic risks would be made by VLOPs and VLOSEs.

94 To draw a distinction that clarifies this point, the situation is different under Article 36 of the DSA where the Commission adopts a decision[96] requiring one or more VLOPs or VLOSEs to take action in order to address a crisis.[97] While Articles 35 and 36 pursue similar objectives, i.e. to require VLOPs and VLOSEs to address risks posed by their services, they function according to different mechanisms that in turn lead to different remedies available to any user that may want to challenge a content moderation decision.

95 In both cases the measure that restricts online content is taken by a VLOP or VLOSE, but Article 36 requires the Commission to adopt a legally binding decision that potentially mandates the implementation of content moderation measures. In this case, the decision of the European Commission could be clearly identified as the source of any interference with the freedom of expression that may be adopted by VLOPs and VLOSEs. This leaves online users who are directly and individually concerned by the decision to bring action before the ECJ and seek its annulment under Article 263(4) of the Treaty on the Functioning of the European Union (''TFEU'')[98], insofar as they have *locus standi* for this action under the ''Plaumann test''.[99]

96 On the contrary, risk mitigation measures adopted under Article 35 of the DSA cannot be linked to a legally binding decision of the European Commission. The Commission participates in shaping private content moderation measures through regulatory dialogue, non-binding guidance and informal discussions. Therefore, there is no act of direct and individual concern against which users can bring action. Any action under Article 263(4) against Article 35 of the DSA is likely to be dismissed, for two reasons. First, the vague wording of Article 35 that does not prescribe any specific interference with freedom of expression.[100] Second, the fact that a reading of the Article does not allow to foresee with sufficient certainty in which specific cases it may require restrictions on the dissemination of legal content.[101]

97 Further to Article 263 of the TFEU, users may also not be able to bring action under Article 265(3) of the TFEU by claiming that the Commission has failed to act and protect the freedom of expression of online users in the exercise of its powers while supervising and orienting the conduct of VLOPs and VLOSEs. There are two reasons to conclude that Article 265(3) of the TFEU is not actionable in this case.

98 First, Article 265 would apply to cases where the EU institutions, bodies, offices and agencies have a clear obligation to take a specific action aimed at ensuring the VLOPs and VLOSEs do not violate the fundamental rights of online users when complying with Article 35 of the DSA. This does not seem to be the case under the DSA as the Commission enjoys

96 See Article 36(1) of the DSA.

97 For the purposes of Article 36 of the DSA, a crisis shall be deemed to have occurred where extraordinary circumstances lead to a serious threat to public security or public health in the Union or in significant parts of it. See Article 36(2) of the DSA.

98 See Article 263(4) of the DSA.

99 As formulated by the ECJ in Case 25/62 *Plaumann & Co v Commission* [1963] ECLI:EU:C:1963:17, p. 107. In this judgement, the ECJ set out the criteria to determine in which cases a natural or legal person can be considered to be ''individually concerned'', which is one of the conditions for *locus standi* under Article 263(4) of the TFEU. In particular, the ECJ held at p. 107 of the judgement that ''*persons other than those to whom a decision is addressed may only claim to be individually concerned if that decision affects them by reason of certain attributes which are peculiar to them or by reason of circumstances in which they are differentiated from all other persons and by virtue of these factors distinguishes them individually just as in the case of the person addressed. In the present case the applicant is affected by the disputed Decision as an importer of clementines, that is to say, by reason of a commercial activity which may at any time be practised by any person and is not therefore such as to distinguish the applicant in relation to the contested Decision as in the case of the addressee*''.

100 In its judgement on the Plaumann case, the ECJ held that ''*persons other than those to whom a decision is addressed may only claim to be individually concerned if that decision affects them by reason of certain attributes which are peculiar to them or by reason of circumstances in which they are differentiated from all other persons and by virtue of these factors distinguishes them individually just as in the case of the person addressed*''. See p. 107 of the OJ publication.

101 The formulation in abstract terms of the obligation in Article 35, and the impossibility to single out affected persons, are factors that render unlikely the fulfilment of the criteria affirmed by the ECJ in Plaumann.

significant discretion in deciding how to exercise its supervisory and enforcement powers and is not required to take specific actions.

99 Second, a natural or legal person can bring action under Article 265(3) where an EU institution, body, office or agency has failed to adopt an act to be addressed specifically to that natural or legal person. This Article is therefore not actionable in instances of failure to protect fundamental rights in a 'mediated' manner through supervision and enforcement over VLOPs and VLOSEs by the Commission.

100 Finally, in addition to actions before the ECJ against the Commission, users would also have no effective redress against the actions of the European Commission as a supervisor and enforcer under the DSA by relying on their rights conferred by the ECHR. The EU is not yet a signatory of the ECHR,[102] therefore proceedings against the European Commission cannot be brought before the ECtHR.

101 The lack of redress channels available to users in such cases appears problematic especially in light of the more 'informal' enforcement history of the European Commission in the past year. For example, in the letter sent by Thierry Breton on 12 August 2024, X was effectively requested to take specific actions in relation to clearly identified content. These actions could result in restrictions on the dissemination of legal but harmful content. This is an episode of 'jawboning' by the European Commission that clearly shows how regulatory expectations can be set without the adoption of acts that can be appealed before a court. While not formalised in an official act, these regulatory expectations can be conducive to concrete restrictions on legal content via the obligations laid down in the DSA. While this can take place across different areas where the European Commission has enforcement powers, such as antitrust enforcement, it presents unique problems under the DSA due to the potential consequences for users' freedom of expression.

102 The obligations that arise under the ECHR, and the jurisdiction of the ECtHR, are limited to its signatories, i.e. the Member States of the Counsil of Europe. The EU shall accede to the ECHR according to article 6(2) of the TEU, but the accession has not yet taken place.

102 In conclusion, users have multiple options for redress against content moderation decisions that violate the terms and conditions they adhered to. However, they have no means to obtain redress against interferences affecting the legal content they disseminate online that are put in place by online intermediaries in pursuit of public policy objectives, and indirectly mandated by legal requirements and regulatory demands. Similarly, VLOPs and VLOSEs would not have standing against disproportionate regulatory demands on the restriction of harmful but legal content, since they are not directly and individually concerned by interferences with the freedom of expression of users.

103 This gap in the redress solutions available to online users is particularly problematic for legal but harmful content, since the source of the interference with this category of content is precisely the regulatory dialogue between the Commission and VLOPs/VLOSEs, where it is determined in which cases the level of systemic risks created by harmful content justifies restrictions on its dissemination.

F. Discussion: Gaps in the Legal Framework to Adress a Hybrid Speech Governance Model

104 Based on the two problems highlighted above, a broader overarching issue can be identified. The new mechanism of public-private cogeneration of policies for the moderation of harmful but legal content in the DSA challenges an approach based on the dichotomy between public and private actors, and the different requirements that apply to them. The constitutional ambiguities of public-private cooperation for online speech moderation have already been discussed in relation to other regulatory schemes and provisions.[103] However, they present peculiar and unique issues under the DSA in relation to the moderation of harmful but legal content. The central role of systemic risk assessment

103 Niva Elkin-Koren, 'Government–Platform Synergy and its Perils' in Edoardo Celeste, Amelie Heldt, Clara Iglesias Keller (eds) *Constitutionalising Social Media* (Oxford: Hart Publishing, 2022); Rocco Bellanova, Marieke de Goede, 'Co-Producing Security: Platform Content Moderation and European Security Integration' [2022] Journal of Common Market Studies 1316.

and mitigation in balancing conflicting interests, and in constituting the basis for an interference with otherwise legal speech, would warrant more safeguards in relation to transparency, accountability and redress for legally mandated interferences with legal speech.

105 This interaction is not captured in EU human rights law, where different obligations are traditionally imposed on public and private actors. This interaction seems to be equally not addressed in the DSA.

106 First, the DSA does not require transparency on the dialogue between the European Commission and supervised intermediaries, nor on how regulatory demands shape private content moderation policies. External observers should be able to clearly understand which private content moderation policies are informed by legal requirements and regulatory demands, and which are merely choices of the intermediary.

107 Second, VLOPs and VLOSEs are under no obligation to indicate in their terms and conditions whether a given restriction is a risk mitigation measure put in place to comply with Article 35 of the DSA. A simple mention in this regard would ensure foreseeability for users of the restrictions stemming from a legal requirement, in line with the conditions in Article 52 of the Charter.

108 Third, despite the numerous procedural and transparency requirements laid down in the DSA for providers of intermediary services, the activities of VLOPs and VLOSEs are not subject to the same constraints to which public actors are, especially as concerns fundamental rights protection. Nonetheless, it would be challenging to identify a clear solution to this shortcoming in the absence of a recognition of full horizontal effects for freedom of expression.

109 In this regard, the question arises as to whether the obligation of Article 14(4) DSA to have due regard for the fundamental rights of the recipients of the service should be interpreted as introducing a direct or indirect horizontal effect of such rights in contractual relationships. Article 14(4) of the DSA, together with Article 5(1) of Regulation (EU) 2021/784, represent an unconventional and innovative legislative technique due to their reference to the Charter to frame the obligations of private entities.[104]

110 The explicit requirement on private actors to respect the fundamental rights of the Charter in the context of their contractual practices is a novelty in EU legislation. If observed through the lens of the conceptual framework on digital constitutionalism, it could be seen as an affirmation of constitutional responsibilities for private actors, with the establishment of a quasi-constitutional framework for content moderation practices. This legislative technique raises several questions on multiple fronts, including on whether the EU has competence to enact rules on fundamental rights protection beyond what is already foreseen in the Charter. In this regard, the question to answer is the meaning that should be ascribed to the fundamental rights obligations of Article 14(4). Article 14(4) operates a vague reference to the fundamental rights of the Charter and does not provide guidance on which could be its legal consequences.[105]

111 For this reason, a clear answer cannot be found in the text alone. Different alternative interpretations have been advanced so far on the meaning of Article 14(4),[106] with three ultimately advocating for the recognition of horizontal direct or indirect effects in connection to the provision.[107] These interpretations

104 Tobias Mast, Christian Ollig, 'The Lazy Legislature. Incorporating and Horizontalising the Charter of Fundamental Rights through Secondary Union Law' (Working Papers of the Hans-Bredow-Institut, Project Results No. 70, 2023), p. 5.

105 Mattias Wendel, 'Taking or Escaping Legislative Responsibility? EU Fundamental Rights and Content Regulation under the DSA', in Antje von Ungern-Sternberg (ed.) *Content Regulation in the European Union* (Trier University and Verein für Recht und Digitalisierung e.V., Institute for Digital Law Trier (IRDT), Volume I, 2023) pp. 81-82; Tobias Mast, Christian Ollig (n 104) p. 1.

106 For an overview of the authors that discussed the interpretation of Article 14(4) of the DSA, see: Tobias Mast, Christian Ollig (n 104); Joao Pedro Quintais, Naomi Appelman, Ronan Fahy, 'Using Terms and Conditions to apply Fundamental Rights to Content Moderation' (2023) German Law Journal 881; Mattias Wendel (n 105).

107 Authors have argued that Article 14(4) could have either a declaratory effect, i.e. merely declaring the horizontal applicability of fundamental rights which stems directly

offer potential solutions to the problem that EU fundamental rights obligations do not fully apply to at least one of the actors involved in the public-private cogeneration of content moderation policies for harmful but legal content.

112 On the one hand, the recognition of the direct horizontal application of freedom of expression would enable to fill a gap in the protection of the freedom of expression of users whose legal but harmful speech is moderated under Article 35 of the DSA. However, further research is needed to operationalise the right to freedom of expression in a horizontal setting, which would prove a difficult task. Due to its traditionally vertical application in binding state action, public law concepts (e.g. legitimacy) would need to be translated to a private setting. Scholars have endeavoured to provide a conceptual framework for the horizontal application of fundamental rights,[108] but it would need to take into account the specificities of each fundamental right in its operationalisation.

113 On the other hand, the indirect horizontal effect of fundamental rights[109] in the contractual relationship between intermediaries and users could at least ensure that contractual provisions are interpreted in light of the Charter. This would enhance the overall level of protection of EU fundamental rights, but it would not lead to a situation where fundamental rights fully constrain the content moderation actions of intermediaries. Thus, a gap would still be left in relation to the moderation of harmful but legal content.

114 Further to Article 14(4), VLOPs and VLOSEs are required to have '*particular consideration*' of the impact of their mitigation measures on fundamental rights, under Article 35(1) of the DSA. This provision is not phrased as laying down a fully-fledged obligation to respect fundamental rights, but rather to take them into consideration in the risk mitigation activities as a procedural requirement.

115 It is unlikely that this provision leads to any horizontal application of fundamental rights that users can rely on, in consideration of both its wording and the observations made above on the hurdles to recognise the horizontal effect of freedom of expression in the EU legal order. The European Commission may rely on this provision to ensure that risk mitigation measures are in line with fundamental rights. However, this does not provide for safeguards against public interferences with freedom of expression, as the European Commission would have the final word.

G. Conclusion

116 Hybrid or meta-regulatory forms of governance have become increasingly popular in EU digital regulation. They present undeniable advantages by giving significant discretion to the same entities that are in the best position to understand and address the risks posed by their services.

117 This contribution does not intend to label these regulatory arrangements as negative or unacceptable, nor to outrightly criticise the DSA. While recognising the positive developments introduced by the DSA, this contribution highlights the preconditions that could, but not necessarily would, allow for non-transparent and unaccountable backdoor entries

from the Charter, or a constitutive effect, i.e. being the source, with constitutive force, of the horizontal effects of the fundamental rights protected by the Charter. For a discussion on the hypothesis that sees Article 14(4) as constitutive of horizontal direct effects, see: Tobias Mast, Christian Ollig (n 104). For a discussion on the hypothesis that sees Article 14(4) as declaratory of pre-existing horizontal direct effects stemming directly from the Charter, see: Mattias Wendel (n 105).

108 David Bilchitz, *Fundamental Rights and the Legal Obligations of Business* (Cambridge University Press 2021).

109 Indirect effect is a doctrine used both in EU institutional law and in national contract law to indicate a situation where a provision or principle has indirect effect because it acts as a source of interpretation of another provision of principle. In EU law, the doctrine of harmonious interpretation (or indirect effect) was developed by the ECJ to require, in certain circumstances, that national law is interpreted in light of EU directives. In the context of contractual relationships, the indirect effect of fundamental rights indicates the role of fundamental rights to act as source of inspiration for interpreting and applying norms of contract law. The indirect effect of fundamental rights in contractual relationships has been mostly discussed at the level of EU Member States' law, especially in German case-law through the doctrine of *mittelbare Drittwirkung*.

of public policy considerations into private policies for the moderation of legal but harmful content. Moreover, it intends to hint at the necessity to discuss possible solutions. As new regulatory models emerge, it is necessary to conceive new solutions to ensure that public functions are performed in a way that is consonant with a democratic system based on the rule of law.