CELE
Centro de Estudios en Libertad de Expresión

Artículo de investigación

64

Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation

Agustina Del Campo, Nicolás Zara, y Ramiro Álvarez Ugarte

February 2025

Agustina Del Campo, Nicolás Zara, y Ramiro Álvarez Ugarte, "Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation", Artículo de investigación No. 64, Centro de Estudios en Libertad de Expresión (CELE), Buenos Aires (2025). Forthcoming in the *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* (vol. 16, No. 2).

UP
Universidad de Palermo

Facultad de Derecho

Centro de Estudios en Libertad de Expresión

CELE
Centro de Estudios en Libertad de Expresión y Acceso a la Información

Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation*

Agustina Del Campo
CELE/UP
adelcampoiso@gmail.com

Nicolás Zara
CELE/UP
nicolaszaracele@gmail.com

Ramiro Álvarez Ugarte
CELE/UP
ramiroau@gmail.com

February 25, 2025

Resumen

This paper discusses the risk-based approach of the Digital Services Act (DSA) of the European Union. By embracing open-ended standards instead of rules and by imposing broad risk-identification and mitigation obligations on private parties, the DSA pushes forward a form of managerial co-regulation that is a paradigmatic shift in platform regulation, that has already influenced other regulatory proposals around the globe. This paper argues that the move is consequential from the perspective of the role of human rights in Internet governance. We posit that the approach pose unique problems when seen from the popular three-prong test used by

* This paper is the result of several months of discussion with the team at CELE. We want to thank Matías González for his insights and valuable feedback on earlier drafts of this work. We also greatly benefited from conversations with Rachel Griffin, Joan Barata, and Alexander Hohlfeld and from our participation in the European Rights & Risks Forum organized by the Global Network Initiative and the Digital Trust and Safety Partnership in June 2024, the Summer Course on European Platform Regulation organized by the University of Amsterdam Law School's Institute for Information Law (IViR) in July 2024, and our involvement in the DSA Civil Society Coordination Group led by CDT Europe. The authors extend special thanks to Professors Aleksandra Kuczerawy (KU Leuven) and Joris van Hoboken (UvA), as well as to the participants in the seminar held at CiTiP (KU Leuven) in December 2024, for their critical perspectives on an earlier version of this paper. Any errors that remain are solely our responsibility.

apex courts around the world to assess restrictions on freedom of expression. Furthermore, we argue that it pushes rights out of the center stage of Internet governance and may create a logic of "symbolic compliance" where governance role of rights is further diminished. Finally, this paper identifies opportunities to address or mitigate the challenges identified, especially in an enforcement stage that remains quite open to these kinds of efforts.

Keywords regulation; Internet governance; Digital Services Act; coregulation; human rights

Introduction

Regulatory proposals for online platforms managing speech online have embraced a risk-based approach that puts the concept of risk at the center stage. The adoption of this approach by important rule-making bodies and organizations in the field of technology is an innovation worthy of study. It happened along two broader paradigmatic shifts in regulatory approaches. First, the tech sector has been slowly but steadily moving from a self-regulatory to a regulatory paradigm;[1] second, trends in regulation have shifted from a "command and control" approach to "new governance" forms of regulation[2] that rely heavily on informal processes of rule-making[3] and an ongoing dialogue between regulators and regulatees.

The European Digital Services Act (DSA)[4] was the first formal regulation to adopt a risk-based approach to speech governance. The Act defines a set of systemic risks that very large online platforms (VLOPs) and search engines (VLOSEs) should assess, including the dissemination of illegal content and other negative impacts of their services on fundamental rights, democratic processes, public health, minors, any person's physical

1 Monroe Price and Stefaan Verhulst, 'The Concept of Self-Regulation and the Internet' in J Waltermann and M Machill (eds), *Protecting our children on the Internet. Towards a new culture of responsibility* (Bertelsmann Foundation Publishers 2000).

2 Julile Cohen and Ari Waldman, 'Introduction: Framing Regulatory Managerialism as an Object of Study and Strategic Displacement' (2023) 86 Law and Contemporary Problems i <https://scholarship.law.duke.edu/lcp/vol86/iss3/10>; Christopher T Marsden, *Internet Co-Regulation. European Law, Regulatory Governance and Legitimacy in Cyberspace* (Cambridge University Press 2011); Chris Marsden, Trisha Meyer and Ian Brown, 'Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?' (2020) 36 Computer Law & Security Review 1 <http://www.sciencedirect.com/science/article/pii/S026736491930384X> accessed 17 August 2020.

3 Robert Gorwa, 'The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content' (2019) 8 Internet Policy Review <https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content> accessed 25 August 2020.

4 European Commission, Digital Services Act 2020 [2020/0361 (COD)].

and mental well-being, and gender-based violence.[5] Platforms are to conduct their own assessments of risks. Upon their findings, they must adopt mitigation strategies, be duly diligent in the measures they adopt, go over yearly external audits that evaluate their compliance,[6] and learn from the process.[7] Furthermore, the DSA empowers regulators to monitor, enforce, and punish these companies for non-compliance or poor compliance through different means.[8] The DSA's approach has since been replicated by UNESCO in their 2023 Platform regulation guidelines,[9] and more recently endorsed by the UN within the Global Digital Compact,[10] the European Union AI Act[11] and, partially, in other pieces of national legislation.[12]

The risk-based approach is not new and can be genealogically traced to previous experiences and regulatory trends. It first emerged in the field of environmental law, consumer protection, and financial services, and it eventually became its own legal "regime"—a particular way to connect and manage certain legal rights and obligations to achieve certain goals. What is unique to these iterations of risk-based approaches is that, when applied to content platforms, they ultimately entail classifying content and speech. And there lies the problem.

Content classification has been a fundamental tool for the protection of freedom of expression under human rights law. International treaties allow only certain types of speech to be legitimately restricted by the state. Clearly determining what content is legal and which is not is key for freedom of expression theory and practice, for the State can only restrict content that is deemed illegal, whether because it infringes upon the rights of others or because it affects an important social interest, as stated, for example, by article 19 of the International Covenant on Civil and Political Rights (ICCPR), article 10 of the European Convention on Human Rights (ECHR), or article 13 of the American Convention on Human Rights (ACHR). State restrictions to freedom of expression need to pass an analysis of legality, legitimate interest, necessity, and proportionality under the most stringent standards. Risk-based regulations applied to content platforms not only allow for but mandate the creation of new categories of speech that fall somewhat short of the binary distinction between what is legal and what is not. They

5 Ibid, pars. 80-84.
6 Ibid, art. 37.
7 Ibid.
8 Ibid 2.
9 UNESCO, 'Guidance for Regulating Digital Platforms. Safeguarding Freedom of Expression and Access to Information Through a Multistakeholder Approach' (UNESO 2023) Final version.
10 GDC, 'Global Digital Compact' (United Nations General Assembly 2024) A/79/L.2.
11 Artificial Intelligence Act 2024 (OJ L).
12 Online Safety Act 2023 2023 (2023 c 50); David Lametti, Online Harms Act (Canada) 2024 [C-63]; Richard Blumenthal and Marsha Blackburn, Kids Online Safety Act 2024 [S. 1409].

also mandate platforms and search engines to address them under the risk identification and mitigation paradigm and expand the universe of content categories to be indirectly governed by the State.

Section two discusses the history of the current trend towards risk-based regulation by highlighting the use of risk as a concept in different legal institutions and the rise, in the 1970s, of a managerial approach to regulation. The DSA combines both. Section three dwells on the reasons why states have turned to a risk-based approach for the tech sector. Section four argues that regulating speech like the DSA does poses unique challenges and trade-offs for freedom of expression. In particular, we discuss the differences between risks under the UNGPs and under the DSA and the tension of the latter approach with the popular three-prong test used by many courts to analyze the legitimacy of restrictions to freedom of expression. We argue that the risk-based approach pushes rights out of the center stage of Internet governance and may create a logic of "symbolic compliance" where their governance role is further diminished. Finally, this paper identifies opportunities to address or mitigate the challenges identified, especially in an enforcement stage that remains quite open to these kinds of efforts.

Risk-based Regulation: A brief history

Risk-based regulation is not new. It has been used in the past in environmental and financial regulations,[13] among others. But its adoption by the European Union in the field of technology and Internet governance[14] has influenced other regulatory proposals at the international,[15] regional[16] and even national levels.[17]

Indeed, risk is a concept used in the law in many ways, sometimes implicitly as a reason to adopt a rule that regulates conduct to prevent an undesirable event from happening or to determine who should carry its costs. This is what happens when the law distributes general duties, like the duty of care or assessment by those who engage in activities deemed risky, or imposes precise rules, like the ones mandating the use of a seat belt while driving or special requirements to transporting hazardous waste. Risk is also present when the law identifies those responsible if legally redressable harms occur.[18] The legis-

13 Cohen and Waldman (n 2).

14 European Commission Digital Services Act (n 4).

15 GDC (n 10).

16 Artificial Intelligence Act.

17 Online Safety Act 2023; David Lametti Online Harms Act (Canada) (n 12); Richard Blumenthal and Marsha Blackburn Kids Online Safety Act (n 12).

18 François Ewald, 'Insurance and Risk' in Graham Burchell, Colin Gordon and Peter Miller (eds), *The Foucault Effect: Studies in Governmentality* (University of Chicago Press 1991) 201 ("Insurance and the law of responsibility are two techniques which bear on the same object").

lator makes a balancing exercise between the costs of risk prevention or risk avoidance and the magnitude of harms to be produced to determine which harms will be legally redressable and which risks of harm are to be managed, for not all harms are to be remedied by law. For instance, the emotional risks involved in engaging in interpersonal relationships, for example, are not redressable by law, but the financial ones sometimes are. Moreover, not all risks are to be managed legally, so you can be a journalist without proper training but you cannot be a lawyer or a doctor without a degree. From a legal standpoint, risk is deeply imbued with normativity.[19]

The law also invokes risks more explicitly as something to be managed and with the purpose of creating value out of things that—without the law—would not have any. As François Ewald put it, from a legal perspective risk is "a specific mode of treatment of certain events capable of happening to a group of individuals—or, more exactly, to values or capitals possessed or represented by a collectivity of individuals: that is to say, a population. Nothing is a risk in itself; there is no risk in reality. But on the other hand, anything can be a risk; it all depends on how one analyzes the danger, considers the event".[20] Identifying a risk is the first step of its operation as a legal mechanism. Once the risk has been named, assessed, and valued, it can be distributed and what previously was a reason not to go somewhere becomes part of the planning process that will take us there. The maritime insurance contract is a good example of this mechanism in action. The possibility of a shipwreck weighs heavily on the mind of a merchant before loading a ship with valuable cargo. What if the ship foundered under an unexpected raging sea? The insurance contract distributes those risks and creates incentives for maritime travel and shipping.[21]

Generally speaking, the way the law operates when dealing with risk is as follows: identification of a risk, of the agent responsible for its management, of the behavior that is expected to be followed or avoided, and of the liability for the eventual harm. This is the essence of how risk operates as a legal mechanism to distribute potential costs implied in human activities and creates incentives for activities deemed beneficial.

Legal systems can identify, assess, and manage risks differently depending on the topic, the complexity of the issues, or the incentives or disincentives they seek to create. Traditional democratic rule-making models heavily support a command-and-control approach, where rules are made through public deliberation and the conduct prohibited or expected is clearly defined and prescribed. In the 1970s, a "new regulatory mood be-

19 Ortwin Renn and Andreas Klinke, 'Risk Governance: Concept and Application to Technological Risk' in Adam Burgess, Alberto Alemanno and Jens Zinn (eds), *Routledge Handbook of Risk Studies* (Routledge 2019).

20 Ewald (n 18) 199.

21 Ibid 199–200.

gan to emerge",[22] that was skeptical of the power of governments alone to identify and assess redressable risks and regulate accordingly. It posited that the state should adopt certain methods and techniques designed to manage processes of capitalist production in its approach to regulation.[23] It rejected the command-and-control model in favor of "relatively informal modes of policymaking and enforcement ... and its emphasis on devolution of regulatory authority to private-sector partners and delegates".[24] Under this approach, which Cohen and Waldman call "regulatory managerialism", general obligations of process are designed by the legislator or administrative body, and the prescribed conduct is replaced with guidances, best practices, compliance certifications, and negotiation.[25]

Those being regulated are an essential part of the processes of regulatory managerialism. Some say the approach is especially suitable for rapidly changing environments: it is "flexible, nimble, responsive to stakeholder priorities, and well suited to a fast-changing, complex economy".[26] It is also good to deal with serious information asymmetries. Those who criticize these techniques find that they are easily co-optable by corporations, who will pursue their own interests at the expense of the public's.[27] They argue corporations can develop check-box compliance approaches that pay lip service to the values being pushed and produce no real change in the world whatsoever.[28] We have recently witnessed the development and expansion of this managerial approach in different spheres of governance, including corporate governance. The state sponsors self-regulatory practices within certain fields, but under its guidance and oversight.[29] It creates obligations upon the private sector to disclose information and seeks to leverage this mandated transparency for different public purposes. Self-regulatory and co-regulatory frameworks encourage the active involvement of those being regulated and evolving obligations and interpretations of the expected conduct along the life of the regulation.[30] Risk is an essential piece of the managerial approach, for it is what authorities can clearly identify and signal as relevant for action, even though they may not have the necessary knowled-

22 Michael Power, *The Audit Society: Rituals of Verification* (Subsequent edition, OUP Oxford 1999) 52.

23 Cohen and Waldman (n 2).

24 Ibid, i.

25 Ibid, iv-v.

26 Ibid, i.

27 Ibid, vi-vii.

28 Lauren B Edelman, *Working Law: Courts, Corporations, and Symbolic Civil Rights* (Illustrated edition, University of Chicago Press 2016).

29 Marsden (n 2); Marsden, Meyer and Brown (n 2).

30 Benoît Frydman, Ludovic Hennebel and Gregory Lewkowicz, 'Co-regulation and the Rule of Law' in Eric Brousseau, Meryem Marzouki and Cécile Méadel (eds), *Governance, Regulation and Powers on the Internet* (Illustrated edition, Cambridge University Press 2012).

ge, information, or incentives to act upon it most efficiently.

I. Technology as a risk to be managed

The extent to which the flow of information on the Internet should be unrestrained or governed, or individual speech acts should be protected or limited, seem to be questions of vital importance in a democratic society committed to values of self-government.[31] Before the DSA, there was no precedent on the application of the managerial model to the field of platform governance through domestic legislation, where either rules of civil liability or immunity laws for third-party posted content were applied.[32] These rules were meant to promote and protect the internet as a means of distribution of speech in a decentralized manner, under the principles of neutrality and non-censorship.[33] But the early immunity approach, designed to encourage the development of an industry unencumbered by potential litigation costs, took some of those key questions out of legal debates. Instead, it created a quid-pro-quo mechanism of paralegal governance that made corporations receptive to government and civil society demands. The state offered immunity but expected collaboration in return.[34] Public officials and civil society organizations (CSOs) became accustomed to asking corporations for actions on a "voluntary" basis. Sometimes, those requests went beyond what the state could accomplish through formal methods of rule-making.

At least since the mid-2010s, immunity rules have come under criticism, and calls to change them have become louder. The growing anxiety regarding the effect of the Internet on democracy seems to have caused the shift. The Internet became a risky thing to be governed, especially since 2016, the year of Brexit, Trump and the Colombian Peace referendum that marked the beginning of the disinformation scare and an era of techlash.[35] Regulatory push-back, such as the Netzwerkdurchsetzungsgesetz in Germany in 2017[36] and the Loi Avia in France in 2020[37] were among the first regulatory efforts against the status quo. The DSA is in many ways a product of these developments.

31 Alexander Meiklejohn, *Free Speech and Its Relation to Self-Government* (Harper & Brothers Publishers 1948).

32 Agustina Del Campo and others, 'Mirando Al Sur. Hacia Nuevos Consensos Regionales En Materia de Responsabilidad de Intermediarios y Moderación de Contenidos En Internet' (AlSur 2021).

33 Communications Decency Act 1996 (USC); Directiva 2000/31/CE 2000.

34 Jeff Kosseff, *The Twenty-Six Words That Created the Internet* (Cornell University Press 2019) 2.

35 Robert D Atkinson and others, 'A Policymaker's Guide to the "Techlash"—What It Is and Why It's a Threat to Growth and Progress' (Information Technology & Innovation Foundation 2019) <https://itif.org/publications/2019/10/28/policymakers-guide-techlash/> accessed 22 August 2022.

36 Netzwerkdurchsetzungsgesetz 2017 (BGBl).

37 Loi No. 2020-766 visant à lutter contre les contenus haineux sur internet 2020 (JORF).

Speech online, however, has always generated concerns that exceed the traditional distinction between legal and illegal content. The volume of content produced daily on the Internet is unprecedented as is the speed at which it spreads, or the fact that, in principle, the content remains up indefinitely. Concerns over potential new harms arising from some of the Internet's structural affordances have been growing louder and have caught the attention of regulators and civil society alike. For instance, can content that would be perfectly legal in isolation become harmful when aggregated? Can its permanence on the Internet be a source of redeemable grievances?.[38] These questions highlight new kinds of potential harms that are not addressed by traditional freedom of expression laws neither locally nor regionally or internationally. Evelyn Douek, for instance, first described the difficulty of assessing a platform's compliance with freedom of expression on a content-by-content basis, as international human rights law traditionally proposes. She posited that, given the volume and scale of content within platforms, content moderation could be assessed in bulk. Compliance with human rights law could be measured based on aggregates where States or companies themselves could determine what percentage of error would be deemed acceptable and the platforms' compliance could be guided by probability and proportionality..[39] She argued that the move was needed to create a content moderation system that was scalable, flexible, adaptable to the ever-changing environment of online speech, and able to treat errors in content moderation decisions as inevitable. More recently, Robert Post has argued that "the scale of the internet produces forms of harm that may best be characterized as stochastic. Previously we asked whether particular speech acts might cause particular harm. The internet has rendered this kind of question almost obsolete. Speech that is simultaneously distributed to billions of persons may produce harm in ways that cannot meaningfully be conceptualized through the lens of discreet causality. We will need instead to think in terms of the statistical probability of harm".[40] He warns, though, that at present "we lack any legal framework capable of assessing stochastic harms in ways that will not drastically

[38] Agustina Del Campo, 'Volume, Speed, and Accessibility as Autonomous Harms: Can Modern Legal Systems Deal With Harmful but Legal Content? - New Digital Dilemmas: Resisting Autocrats, Navigating Geopolitics, Confronting Platforms - Carnegie Endowment for International Peace' (*Carnegie Endowment for Democracy*, 29 November 2023) <https://carnegieendowment.org/2023/11/29/volume-speed-and-accessibility-as-autonomous-harms-can-modern-legal-systems-deal-with-harmful-but-legal-content-pub-91082> accessed 11 January 2024.

[39] Evelyn Douek, 'Governing Online Speech: From "Posts-As-Trumps" to Proportionality and Probability' [2020] SSRN Electronic Journal <https://www.ssrn.com/abstract=3679607> accessed 23 April 2022.

[40] Robert Post, 'The Internet, Democracy and Misinformation' 8 <https://papers.ssrn.com/abstract=4545891> accessed 4 November 2024.

over-regulate speech".[41] In many ways, the risk-based approach adopted in the DSA, and other provisions since, have built their legitimacy on the need to address these new harms that technology generates and offer a path forward.

The risk-based approach now adopted by formal regulation was previously pushed on corporations through the voluntary and soft law approach of the UN Guiding Principles (UNGPs) on Business and Human Rights. This model was supposed to "internalize" corporate commitments to human rights,[42] that could not be imposed externally through hard law because of the gridlock affecting the UN on the issue of human rights and transnational corporations. Risk played a meaningful role in its design. Under the UNGPs, corporations are committed to identifying risks to human rights, and monitoring and evaluating their actions[43] to take "adequate measures for their prevention, mitigation and, where appropriate, remediation".[44] To this end, corporations resorted to processes of impact assessments, procedural devices designed to help them make better decisions regarding their operations and their impact that were particularly well known in the environmental field of law,[45] and applied it to human rights.[46] A whole industry of consultants, experts, and knowledge emerged as a consequence. The field of Business and Human Rights produced the professional cadre that regulatory managerialism needed to operate in the content platform industry.

In sum, the new generation of platform regulation embraces a managerial, co-regulatory model, where risk plays the fundamental role of bridging the gap between state desires to deal with certain harms—some poorly identified and currently not legally redressable—and the much-needed collaboration of those corporations in the position to address them. In this evolution, the identification of the Internet as a risk plays a crucial explana-

41 Ibid.

42 James Harrison, 'Human Rights Measurement: Reflections on the Current Practice and Future Potential of Human Rights Impact Assessment' (2011) 3 Journal of Human Rights Practice 162, 108 <https://academic.oup.com/jhrp/article/3/2/162/2188745> accessed 14 May 2020.

43 John Ruggie, 'Protect, Respect and Remedy: A Framework for Business and Human Rights' (Human Rights Council eport of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises 2008) A/HRC/8/5, par. 25.

44 John Ruggie, 'Guiding Principles on Business and Human Rights. Implementing the United Nations "Protect, Respect and Remedy" Framework' (Human Rights Council eport of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises 2011) HR/PUB/11/04 principle11.

45 John Glasson, *Introduction to Environmental Impact Assessment* (Taylor & Francis Ltd 1998).

46 Desirée Abrahams and others, 'Guide to Human Rights Impact Assessment and Management' (International Business Leaders Forum, International Finance Corporation & el Global Compact de las Naciones Unidas 2010).

tory role, and the path-dependency of the informal mechanisms of governance allowed by the old rules on intermediary liability and the UNGP framework explain the managerial and procedural turn as something more like an evolution than a clean break with the past.

The risks of the risk-based approach as applied to speech

Although risk-based approaches are not new and the risk-based model borrows language and processes from existing human rights soft law documents, the risk management system in the DSA poses new challenges. The DSA broadens the scope of the risks it mandates corporations to mitigate as compared to the UNGPs. Under the DSA, the violation of fundamental rights is only one risk to be addressed among many. Furthermore, the DSA adopts a hard law approach that invokes the coercive power of the state. Because it is tailored to content platforms, it deals mainly with third-party posted content and thus must be scrutinized under freedom of expression standards. Finally, the DSA also expands the speech to be governed by imposing obligations on corporations to act on speech that is, according to standard human rights principles, out of State action reach. As a result of these factors combined, human rights lose centrality and fade into the background of the DSA's risk-based approach.

Differences in kind between UNGPS and the DSA

Unlike the UNGPs, which focus exclusively on risks related to human rights, the DSA treats adverse effects on human rights as just one risk among many.[47] This outward expansion is consequential, for human rights are a framework that plays a somewhat constraining function. The UNGPs don't target any harmful conduct that companies may produce, but only those that may infringe upon human rights. Human rights law, for example, mandates that certain harms be tolerated in democratic societies. Therefore, not every infringement upon the right to privacy or the right to honor, for instance, may be legally redressable. Restrictions to freedom of expression are illegitimate unless necessary, proportionate, and well-defined in the law in order to be legitimate, even if the expression in question might have produced harm.[48] This standard test has been es-

[47] Rachel Griffin, 'What Do We Talk about When We Talk about Risk? Risk Politics in the EU's Digital Services Act - DSA Observatory' (*DSA Observatory*, 31 July 2024) <https://dsa-observatory.eu/2024/07/31/what-do-we-talk-about-when-we-talk-about-risk-risk-politics-in-the-eus-digital-services-act/> accessed 5 November 2024; European Commission Digital Services Act (n 4), article 34.1(b).

[48] CIDH, 'Marco Jurídico Interamericano Del Derecho a La Libertad de Expresión' (Relatoría Especial para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos 2009)

tablished through a shared practice that has produced a corpus of standards, case law, precedents, and rules that defines and distinguishes legally redressable from non legally redressable harms and, therefore, limits the kinds of risks of harm that companies need to address under these norms. The expansion of "risks" that the DSA encourages is less constraining, for the harms to be mitigated are more vague, are not attached to any particular legal framework, and are built on less developed foundations. The law, for example, requires auditors to have expertise in risk management in general;[49] expertise regarding "the systemic societal risks referred to in Article 34" is also expected.[50] Notably, but not surprisingly no expertise is required in the field of human rights. This is especially telling given the lack of standards or benchmarks provided by the delegated act on independent audits,[51] the centrality of auditors in the DSA's oversight infrastructure,[52] and how audit results can inform regulatory supervision.[53]

The UNGPs are a soft law instrument that companies may voluntarily choose to abide by and that lack formal enforcement mechanisms. They came to be as a way of dealing with the thorny question of business and human rights inside the United Nations, crossed by a pervasive disagreement between the countries that produced transnational corporations (in the North) and those who received them (in the South). The UNGPs were the way out of that gridlock, and they were meant to deal with resource-intensive industries with profound social and environmental impact on the ground, such as the extractive industries.[54] They assumed that the main risk for human rights came from states, but transnational corporations could on occasion violate them or contribute to their violation. They were meant to close the governance gap "created by globalization—between the scope and impact of economic forces and actors, and the capacity of societies to manage their adverse consequences. These governance gaps provide the permissive en-

OEA/Ser.L/V/II CIDH/RELE/INF. 2/09, par. 67.

49 European Commission Digital Services Act (n 4), article 37.3.

50 Commission Delegated Regulation (EU) 2024/436 of 20 October 2023 supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council, by laying down rules on the performance of audits for very large online platforms and very large online search engines 2023, recital 9.

51 Ibid.

52 Giovanni De Gregorio and Oreste Pollicino, 'Auditing Platforms under the Digital Services Act' [2024] Verfassungsblog <https://verfassungsblog.de/dsa-auditors-content-moderation-platform-regulation/> accessed 25 November 2024.

53 Commission Delegated Regulation (EU) 2024/436 of 20 October 2023 supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council, by laying down rules on the performance of audits for very large online platforms and very large online search engines, recital 1.

54 Ramiro Álvarez Ugarte and Laura Krauer, 'ICT and Human Rights: Towards a Conceptual Framework of Human Rights Impact Assessments' (Centro de Estudios para la Libertad de Expresión 2020).

vironment for wrongful acts by companies of all kinds without adequate sanctioning or reparation".[55] The DSA batters on a similar nail through a much more powerful hammer. Unlike the UNGPs, the DSA is hard law and foresees enforcement mechanisms, penalties, and sanctions against companies that fail to comply with its mandates. The difference becomes significant particularly when dealing with content platforms and an open framework of "risks" as discussed above.

Finally, the DSA is tailored to deal with content-facilitating and content-producing companies within the Internet's decentralized architecture. It targets a particular industry and explicitly articulates the new paradigm: that online speech generates risks that need to be managed, mitigated, or else. And while the UNGPs are focused on human rights obligations, the narrative[56] the DSA tells is not that of rights but one centered on risks: in its risks and harms approach, human rights concerns are present, but only as one risk among many. And even when the DSA does have Human Rights safeguards built into its text, if not taken seriously, they could even end up legitimizing state action in violation of Human Rights (for instance, the insufficient safeguards in article 9 against illegitimate state orders).

Furthermore, the DSA fails to acknowledge the state as a risk for the protection of human rights. Mandated transparency is among the most celebrated measures adopted by the DSA, as is data access for researchers, but every procedural measure incorporated in the DSA is directly tied to what is considered relevant to address the risks identified in Article 34. There are ongoing discussions about the possibility of State agents engaging with companies as trusted flaggers and no duty to report on the potential correlations this may bring about. The framework is prone to obscure rather than provide transparency to state-led censorship. At the center of the DSA lays a right deemed fundamental for the working of democratic societies: the right to freedom of expression. Unlike the extractive industries contaminating the environment or exercising violence upon local populations, the transnational corporations targeted by the DSA are in the business of facilitating communications between individuals. It is an economic activity, but one closely connected to the exercise of the fundamental rights of their users.

Expanding the speech to be governed

The correlation between the expressions generally frowned upon by society and those considered illegal under the law is not perfect. As a general rule, under either international human rights law, constitutional law, or both, all speech is protected under the right to freedom of expression, with very limited exceptions. This leaves a lot of offensive, un-

55 Ruggie (n 43), par. 3.

56 Robert Cover, 'Foreword: Nomos and Narrative' (1983) 97 Harvard Law Review 4.

pleasant, shocking expressions still protected by the law. Those expressions are generally referred to as "lawful but awful" or "harmful but legal" speech —new categories that grew out of the anxieties produced by the effects of certain Internet-based speech on democratic societies. The risk-based approach that the DSA embraces allows it to expand its governance over these new categories of speech.

Under old-school intermediary liability laws, each platform had the discretion and the incentive to address the problem of "harmful but legal" content as they saw fit through enforcement of their terms of service, without concerning themselves with potential liability stemming from under or over-removal of content. This structure enthroned platforms as digital sovereigns, private censors, and new governors of speech.[57] The DSA challenges this model and brings about a system of content governance to tackle the issues posed by lawful but awful expression online.

The risk-based approach is presented as useful for this venture because it allows the state to shape content moderation practices by intervening mostly on processes, which, in turn, awards each platform the much-needed flexibility to tailor their interventions to the very specific risks derived from their operations. In the DSA's model, it is not the legislator who identifies the kind of legal but harmful content they want platforms to disallow. Such a law would most probably violate freedom of expression guarantees. Platforms are the ones that must identify, assess, and mitigate the specific risks that their affordances or the use of their services generate. However, the risks that platforms should look for are outlined by the legislator, mostly on Article 34. Additionally, the DSA prescribes some of the measures that could be taken to mitigate those risks.[58] While there are no explicit mandates to remove certain categories of content, the expansion of the risks that companies should mitigate beyond those identified in international human rights law per se expands the categories of speech governed. The list is not exhaustive, so companies are permitted to engage in different, innovative mitigation measures exceeding those included in the law.

Rather than imposing hard metrics or targets, the DSA seeks to encourage platforms and search engines to "think" about the risks they potentially generate.[59] It presents itself as a holistic approach that stays away from drawing clear lines or establishing clear-cut

57 Kate Klonick, 'The New Governors: The People, Rules, And Processes Governing Online Speech' (2018) 131 Harvard Law Review 73; Jack M Balkin, 'Old-School/New-School Speech Regulation' (2014) 127 Harvard Law Review 2296 <https://www.jstor.org/stable/23742038> accessed 10 March 2023.

58 European Commission Digital Services Act (n 4), article 35.

59 Evelyn Douek, 'The Siren Call Of Content Moderation Formalism' in Lee Bollinger and Geofrey Stone (eds), *Social Media, Freedom of Speech, and the Future of our Democracy* (Oxford University Press 2022).

rules that mandate the removal of content other than the illegal one. However, although not prescriptively, article 35 suggests the expeditious removal of hate speech and cyber violence as an effective and desirable mitigation measure.[60] The Codes of Conduct on Disinformation and Hate Speech also provide concrete suggestions like demonetization, filtering, blocking or deindexing for harmful but legal content. By maintaining platform immunity for third-party posted content, the DSA keeps companies safe from the liability arising from their own "errors" in content moderation.

The DSA has been portrayed as a law mainly dealing with processes. It mandates companies to offer a series of appeals systems and complaint mechanisms designed to make sure that content producers' and their audiences' rights are respected and terms of service are applied to them consistently.[61] A series of information disclosure obligations, such as transparency reports, independent audits, and data access for researchers, are incorporated into the law to make platform accountability possible. The inclusion of these procedural obligations upon companies is a significant and well-received contribution to platform governance debates.

However, the DSA does not only regulate processes but also deals with substantive issues. The systemic risks it identifies force corporations to assess the nature of content and act upon legally protected speech. It also mandates companies to assess their own tools and means to distribute and organize legally protected speech. Some may argue that the DSA only generates obligations vis-à-vis the companies' own actions rather than those of content producers, but VLOPs and VLOSEs systems are directly tailored to present, distribute, and curate content. And both the generation and distribution of content are essential parts of well-established freedom of expression laws and standards all over the world. In well-functioning democracies, limiting the reach of a newspaper by fixing its selling price through law or limiting the number of copies that may be printed of a given book or magazine would be as unconstitutional as censoring it.

Systemic risks and compliance with freedom of expression human rights standards

Voluntariness matters when non-state actors are involved, especially when it comes to speech. For instance, a microblogging platform for and by puppy owners could establish an "only pictures of puppies" rule without triggering human rights concerns. Ho-

60 European Commission Digital Services Act (n 4), article 35.

61 Pietro Ortolani, '"If You Build It, They Will Come". The DSA "Procedure Before Substance" Approach' in Joris van Hoboken and others (eds), *Putting the Digital Services Act Into Practice: Enforcement, Access to Justice, and Global Implications* (1st edn, Verfassungsblog gGmbH 2023); European Commission Digital Services Act (n 4), article 21.

wever, if state-mandated, the rule would trigger heightened human rights-based scrutiny as a potential infringement on freedom of expression. The nature of the right affected is also important. Patrimonial rights, for instance, are generally easier to limit than those deemed essential for the good working of democratic institutions. Constitutional courts have generally shown more deference to legislators affecting the former rather than the latter.

The DSA seeks to navigate these important distinctions. On the one hand, it grants corporations a lot of leeway to manage their services as they see fit—they just need to be aware of, and manage, a set of very vague risks that the European legislator identified as "systemic". It does not regulate the patrimonial rights of corporations (that's a job for the Digital Markets Act) but it claims not to regulate speech rights either. The DSA, we are told, is about processes rather than substance. But the nature of the companies being regulated pulls speech rights back in, for issues as varied as disinformation and misinformation, the sale of illegal products, online scams, election interference, hate speech and discrimination, and terrorism-promotion content, all include, quite obviously, a freedom of expression dimension that demands that we carefully assess whether restrictions based on the state's interest to combat these harms are necessary in a democratic society.[62] If corporate action upon third-party posted content is directly linked to a state mandate, such restriction should be subjected to the scrutiny called for by the three-prong test. And the distance between the action and what was required—a distance established by design—does not fare well under this light. Article 53(3) of the European Charter, which is directly cited by the DSA, mandates that the rights contained therein be interpreted in light of the European Convention on Human Rights.

0.0.1. The legality principle Any restriction to freedom of expression must be prescribed by law. Under stable criteria of the ECtHR, any restriction to freedom of expression must be "formulated with sufficient precision to enable the citizen to regulate his conduct: he must be able—if need be with appropriate advice—to foresee, to a degree that is reasonable in the circumstances, the consequences which a given action may entail".[63] The ICCPR adopts a similar standard..[64] Laws restricting freedom of expression "may not confer unfettered discretion on those charged with their execution".[65]

62 Tarlach McGonagle and Onur Andreotti, *Freedom of Expression and Defamation* (Council of Europe 2016) 12.

63 *The Sunday Times v the United Kingdom (no 1)* [1979] ECtHR 6538/74, par. 49.

64 HRC, 'General Comment No. 34 on Article 19 of the ICCPR' (Human Rights Council 2011) CCPR/C/GC/34, par. 25.

65 Ibid; HRC, 'General Comment No. 27 on Article 12 of the ICCPR' (Human Rights Council 1999) CCPR/C/21/Rev.1/Add.9, part. 13.

Absolute legal precision is not, however, the standard—"experience shows this to be unattainable", and "whilst certainty is highly desirable, it may bring in its train excessive rigidity and the law must be able to keep pace with changing circumstances. Accordingly, many laws are inevitably couched in terms which, to a greater or lesser extent, are vague and whose interpretation and application are questions of practice".[66] As it stands today, and without proper guidance from the European Commission, it remains unclear which is the risky content that platforms and search engines need to identify and mitigate to fulfill their Article 34 and Article 35 obligations under the DSA.

Proponents of the risks-based approach in the DSA distinguish the clarity needed to directly restrict speech from that required to hold companies accountable for speech-generated harms. They argue that vagueness in the categories of risks listed in Article 34 and the lack of concrete definitions of risks might be features, rather than bugs, of the risk-based approach in the DSA.[67] They introduce a degree of flexibility that allows both VLOPs and VLOSEs and the European Commission to initiate an iterative process where they can jointly set goals according to their growing capabilities and build on previous findings. This, they hope, will encourage a kind of regulatory dialogue where progress is made developmentally. They favor, thus, a flexible framework that would be more responsive to the fast-paced and ever-changing activity of the sector, which could render more rigid systems with static rules and bright lines useless.[68]

However, the DSA is concerned with harms caused by categories of speech defined in the broadest terms and distributed, organized and curated by companies. As long as there are people expressing themselves, there will be risks to the well-being, health, security, and even the enjoyment of some human rights as conceived in the DSA. Platforms cannot completely mitigate these risks without shutting down their operations entirely. So difficult questions come up. How much of each risk could reasonably remain unmitigated and what are the relevant metrics to be used for each? Are metrics as effective to encapsulate these risks as those used to measure water's fitness for human consumption? Human expression is inherently complex—tensions will appear regarding the proper balancing between freedom of expression vis-à-vis countervailing interests, protection against certain risks, and even frustration of the human rights of third parties,

66 *The Sunday Times v. the United Kingdom (no. 1)* (n 63), par. 49.

67 Zohar Efroni, 'The Digital Services Act: Risk-Based Regulation of Online Platforms' [2021] *Internet Policy Review* <https://policyreview.info/articles/news/digital-services-act-risk-based-regulation-online-platforms/1606> accessed 24 November 2024.

68 Justin Hurwitz, 'Regulation as Partnership' (2019) 3 Journal of Law and Innovation 1; Tim Wu, 'Agency Threats' (2011) 60 Duke Law Journal 1841 <https://heinonline.org/HOL/Page?handle=hein.journals/duklr60&id=1857&div=&collection=>; Stuart Brotman, 'Communications Policy-Making at the FCC: Past Practices, Future Direction' (1988) 7 Cardozo Arts & Ent LJ 55.

or between different interpretations of freedom of expression, the adoption of either of which would lead to inevitably different outcomes.[69]

The "systemic" component of risks as early as we are in the implementation of the DSA poses additional challenges. It is still unclear whether the systemic aspects refer to the systems within a single company, a group of similar companies that together create a system, a larger group of companies encompassing hardware and software providers and their consumers, or an even broader interpretation—as in the system formed by the outlets and the agents whose interaction form a Habermasian public sphere.[70]

Looking to the financial services literature, where the idea of "systemic risk" comes from, is not helpful (nor is the inspiration promising, considering the success of financial regulation to prevent abuses and harms). Broughton Micova and Calef, after looking at financial markets, suggested that "the systemic nature of risk is not only about the number of users affected by any harm but also derives from the way very large services function as public spaces and from the potential for effects on public systems due to the scale and role of the services designated as VLOPs and VLOSEs".[71] And yet, this concept remains a "remarkably broad category" in the context of the DSA.[72] Article 34 does not offer a definition and researchers disagree.[73] As Griffin puts it, "there are deep ideological and political conflicts over the nature of these essentially contested concepts".[74] The hard work of defining "risk area-specific understandings of what systemic failure or crisis looks like and what effects contribute to those" remains ahead of us.[75]

While there are visible advantages to a flexible approach towards evolving technolo-

---

69 Ramiro Álvarez Ugarte, 'From Soft Law to Hard Law: Human Rights Impact Assessments in the Digital Services Act Era | TechPolicy.Press' (*Tech Policy Press*, 20 June 2024) <https://techpolicy.press/from-soft-law-to-hard-law-human-rights-impact-assessments-in-the-digital-services-act-era> accessed 29 October 2024.

70 Jürgen Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (The MIT Press 1991).

71 Sally Broughton Micova and Andrea Calef, 'Elements for Effective Systemic Risk Assessment Under the DSA' [2023] SSRN Electronic Journal 49 <https://www.ssrn.com/abstract=4512640> accessed 25 November 2024.

72 Paddy Leerssen, 'Outside the Black Box: From Algorithmic Transparency to Platform Observability in the Digital Services Act' (2024) 4 Weizenbaum Journal of the Digital Society 24 <https://ojs.weizenbaum-institut.de/index.php/wjds/article/view/4_2_3> accessed 3 August 2024.

73 Oliver Marsh, 'Researching Systemic Risks Under the Digital Services Act' (*Algorithm Watch*, 26 July 2024) <https://algorithmwatch.org/en/researching-systemic-risks-under-the-digital-services-act/> accessed 5 November 2024.

74 Griffin (n 47).

75 Broughton Micova and Calef (n 71) 50.

gies, the vagueness of the existing categories gives both companies and the European Commission a great deal of discretion, which is exactly what the principle of legality was meant to prevent. There is, then, a fundamental tension between the risk-based approach as broadly defined in the DSA and the black-lettered stable rules established by the European Court of Human Rights to assess restrictions on freedom of expression. This tension will have to be resolved in the future, either by insisting on the need for clarity and precision or by relaxing the legality principle.

The legitimate aim This is probably the easier part of the three-prong test for the DSA to pass. International Human Rights Law requires that limitations be justified to pursue a legitimate aim and sets out what those legitimate objectives may be. The ECtHR has consistently held that there is little scope under Article 10.2 for restrictions on political speech or the debate of issues of public interest.[76] However, the Court has also been deferential to the kind of arguments that states usually deploy to justify restrictive measures. The requirement that the state justifies them under a "legitimate objective" has often been easily met.[77]

The harms derived from speech are actual, real, and in many cases serious. The impact of technology on public discourse and society as a whole is undeniably deserving of attention. Tech companies need to be held accountable for the legally redressable harm they produce or contribute to producing. States are allowed to address them and restrict freedom of expression when necessary but, in that venture, they are also obligated to respect this three-prong test. In its own problematic way (as discussed in the previous section), the DSA has invoked a set of aims that are indeed legitimate. So we move on quickly to the next step of the analysis.

The necessity and proportionality test To be compatible with the ECHR, restrictions must be "necessary in a democratic society" and proportional, which means that they must correspond to a pressing social need and be proportionate to the legitimate aims being pursued.[78] Under the ECHR, proportionality requires that restrictions are adopted through the least restrictive means to achieve the goals pursued by the regulation.[79] Under the ICCPR, proportionality requires measures not to be overbroad and

[76] *Castells v España* [1992] Tribunal Europeo de Derechos Humanos 11798/85; *Wingrove v United Kingdom* [1996] European Court of Human Rights 17419/90, HUDOC.

[77] Lorna Woods, 'Freedom of Expression in the European Union' (2006) 12 European Public Law 371, 376 <https://kluwerlawonline.com/journalarticle/European+Public+Law/12.3/EURO2006026> accessed 5 November 2024.

[78] McGonagle and Andreotti (n 62) 12.

[79] *Axel Springer Se And Rtl Television Gmbh V Germany* [2017] European Court of Human Rights 51405/12, HUDOC; *Perinpçek v Switzerland [GC]* [2015] European Court of Human Rights

"the least intrusive instrument amongst those which might achieve their protective function".[80] They must be "proportionate to the interest to be protected".[81] The principle must be respected both in the law establishing the restriction and in the specific instances in which it is enforced.[82]

The DSA mandates that risk assessments should be "proportionate to the systemic risks, taking into consideration their severity and probability".[83] And mitigation measures should also be "proportionate".[84] However, when compared to the international standards reviewed, the DSA framework tweaks proportionality in some relevant ways. First, by requiring that the obligations under Articles 34 and 35 are complied with in a way that is proportionate to the risks identified and reported, the DSA is departing from the understanding of proportionality under international human rights and European fundamental rights law. Under international standards, proportionality contains an objective dimension, one that requires using the least restrictive means possible on each occasion. Under Article 35 of the DSA, mitigation measures must be "proportionate" to risks. Hence, the bigger the (self-assessed) risk, the more stringent acceptable mitigation measures can be. Second, the DSA outsources the determination of the proportionality of mitigation measures to VLOPS/VLOSEs with no further guidance than stating—redundantly—that they must fulfill their assessment and mitigation obligations in a way that is proportionate to the risks identified. Third, instead of assessing the proportionality of each decision where expression is affected (each modification in the terms of service, each removal, each demotion, etc), it looks at platforms' conduct on an aggregate basis. This makes the proportionality of platform content moderation decisions very hard to oversee outside of the internal complaint mechanisms and out-of-court dispute settlement entities provided for in Articles 20 and 21.[85] The instructions set out in articles 13 and 14 of the Delegated Act on independent audits are written in broad terms and lack the granularity and nuances necessary to evaluate proportionality. Although the instructions mandate that auditors evaluate proportionality, reasonableness, and effectiveness, the indicators identified only address the latter. There are no proposals for indicators to measure proportionality or reasonableness vis-à-vis other human rights. This lack of concrete indicators allows auditors to work with radically different benchmarks, so the performance evaluations of different VLOPs or VLOSEs might not be comparable, even when some of them could be to some extent similar in their affordan-

27510/08, HUDOC.

80 HRC (n 64), par. 34.

81 Ibid, par. 34.

82 Ibid, par. 34.

83 European Commission Digital Services Act (n 4), article 34.

84 Ibid, article 35.

85 Ibid, articles 20 and 21.

ces.

To be fair, setting a priori benchmarks as a one-size-fits-all solution would not necessarily be a better idea, for it would impair the auditors' ability to contemplate the risks inherent to platforms of different natures. It makes sense to settle for more realistic expectations in connection with the kind of control that auditors can exert over platforms. Perhaps audits are useful in determining only one aspect of proportionality: whether the narrative presented by platforms makes sense, that is, whether the chosen measures are effective against the specific risk they were intended to mitigate and "reflect the severity of the risk to society and platform users identified by the platform".[86] However, the auditing process seems ill-suited to assess whether the measures were among the least restrictive means to pursue the same policy goals. This is aggravated by the fact that compliance with the voluntary codes of conduct is also evaluated in the audits, as per Article 37 (1)(b).

The incentives in the risk management system of the DSA

Every legislation creates incentives. The immunity laws that characterized the prior generation of platform regulations created incentives for self-regulation and the development of increasingly complex terms of service and content moderation practices and techniques. The DSA creates some problematic incentives vis-à-vis content moderation that we have already developed. But it can potentially have other impacts that may be as problematic and need to be addressed and monitored closely during implementation. The risk-based approach of the DSA is not immune to the critiques towards similar procedural regulation, which in many cases have failed to bring about real change and have instead generated disappointment..[87] We would like to focus on two issues: the decentering of rights produced by the risk-based approach and the risk of "symbolic compliance".

Rights fade into the background Behind any state decision regarding a legal determination of risk "is the question of what is safe enough, implying a normative or moral judgment about acceptability of risk and the tolerable burden that risk producers can impose on others".[88] How States assess this threshold of tolerance "provide hints over what kind of mental images are present and which moral judgments guide people's

[86] Jeff Allen and Abagail Lawson, 'On Risk Assessment and Mitigation for Algorithmic Systems' (Integrity Institute 2024) 52–53 <https://integrityinstitute.org/news/institute-news/risk-assessment> accessed 25 November 2024.

[87] Nikolas Rose and Peter Miller, 'Political Power Beyond the State: Problematics of Government' (1992) 43 The British Journal of Sociology 173 <https://www.jstor.org/stable/591464> accessed 18 March 2024.

[88] Renn and Klinke (n 19) 209.

perceptions and choices" in that state.[89] Reading articles 34 and 35 of the DSA under this lens allows us to understand the relative importance awarded to fundamental and human rights vis-à-vis other risks brought about by big internet platforms. We believe the outcome to be unsatisfactory.

Indeed, the DSA changes the framing of platform regulation. The risk identification, assessment, and mitigation mandates within the DSA seek to strike a balance between competing and often contradicting interests at play in platform governance—innovation, commercial interests, protection of rights, and protection of state interests. It adopts a flexible stand in order to allow for different kinds of platforms to operate, and different approaches and business models to flourish. The premise of regulatory managerialism is that companies have the expertise and the knowledge that regulators lack. It is this information asymmetry that normative flexibility is meant to overcome.

This reframing, however, does not exempt the resulting rules from human and fundamental rights scrutiny. De Gregorio holds that the risk-based approach championed by the EU in the DSA (also in the GDPR and the AI Act) not only can coexist but is also intimately connected to a rights-based approach.[90] However, under this interpretation risks would take the place formerly occupied by rights as the objective parameter against which all the other elements of legislation are measured. It pushes rights to a subservient place—they become another risk category as Article 34 shows. Rights assume a new role— they are "embedded" in the risk analysis and can be "managed" or measured using the same analytic categories.

This shift entails a realignment in policy priorities. The risk-based approach of the DSA downplays human rights analysis and replaces the proportionality analysis that tests state regulations against strict requirements of justification for a risk assessment process where rights are another interest to be balanced against competing interests and concerns. In this exercise, rights no longer hold a preferred position. With big fines looming on the horizon,[91] companies have incentives under the DSA to err on the side of over-mitigation—they simply cannot afford to leave risks insufficiently mitigated and be found noncompliant. To put it in risk management language, the legal risks of under-mitigation are way higher than the alternative. It seems safer to overstate risks than to overstate rights. Auditors will not be of much help, because it is unlikely they will have the necessary information to second-guess companies in their own risk-assessments.

89 Ibid.

90 Giovanni De Gregorio, 'How Does Digital Constitutionalism Reframe the Discourse on Rights and Powers?' (*Ada Lovelace Institute*, 7 December 2022) <https://www.adalovelaceinstitute.org/blog/digital-constitutionalism-rights-powers/> accessed 12 November 2024.

91 European Commission Digital Services Act (n 4), articles 74, 76.

Audits might only show us whether the mitigation measures taken by companies are efficient in tackling the risks stated in their risk reports. We have not seen any investigation or RFIs open on the basis that a company went too far in protecting public health or civic discourse, and we don't envision any either. Ultimately, more restrictive terms and conditions can be attributed to companies' stricter policies, and being private and voluntary, these can be deemed independent from their legal obligations. Regulators hold some power to prevent this from happening. If the DSA does propose a regulatory dialogue moving forward, regulators might push back against companies that overestimate risks. This is not an entirely impossible scenario, but it seems unlikely considering the political and ideological drivers currently in place, including the centrality of risks and the displacement of rights, the absence of the state as a potentially threatening actor in the DSA landscape, and the lack of concern for over-removal. These factors limit potential remedies to this problem to the appeals systems in the platforms, out-of-court settlement dispute mechanisms and private enforcement of the DSA, driven by individual content producers or their audiences.

Rights as checkboxes The trends towards "managerialization" and "proceduralization" of human rights through regulations that establish mandatory due diligence obligations give rise to novel challenges. One of them is linked to the UNGPs as the precedent to the DSA, an approach that has proven limited and that introduces a "distortion" in the very idea of human rights as legal institutions. In many ways, human rights in the UNGPs framework are deprived of some of their essentially legal features such as "enforcement mechanisms, liability, and penalties".[92] While the DSA brings the law back in, the gaps identified previously endure.

One of the challenges that remain ahead, and that those in charge of implementing the DSA should carefully consider, is the risk that Laura Edelman called of "symbolic compliance"—when rights acknowledged in laws and other regulations are taken by corporations as opportunities to develop ritualistic but rather ineffective measures, such as assigning resources, cresting positions, and developing procedures that ultimately fail to produce meaningful conduct change.[93] In these processes, certain actors within corporations gain power but have a limited impact on corporate decision-making, at best.[94]

[92] Ramiro Álvarez Ugarte, 'Bad Cover Versions of Law. Inescapable Challenges and Some Opportunities for Measuring Human Rights Impacts of Corporate Conduct in the ICT Sector' (2024) preprint, under review.

[93] Edelman (n 28).

[94] John W Meyer and Brian Rowan, 'Institutionalized Organizations: Formal Structure as Myth and Ceremony' (1977) 83 American Journal of Sociology 340 <https://www.journals.uchicago.edu/doi/10.1086/226550> accessed 10 September 2024; Tricia Olsen and others,

The fact that risk management approaches are becoming, under the weight of the DSA, a form of hard law creates a number of questions that will remain open as organizations comply and adapt to the regulatory dialogue the act promises. For instance, it is likely that the flexibility of risk assessments under the UNGPs will be lost under the pressure of actual regulations that companies will have to comply with (and it will make no sense for companies to develop different risk assessment procedures). Black-letter law will take over voluntary corporate practices.[95] Many companies that developed APIs to encourage developers services, for example, closed them down under the weight of the GDPR. Would something similar happen under the DSA? It is also likely that corporations will develop positions and adapt their structure to the new laws, and work under a paradigm of compliance,[96] that could be, however, symbolic if previous research is ascribed with a predictive function. This will turn human rights into corporate checkboxes to be filled in a compliance exercise. The fact that they appear somewhat secondary to the primacy of risks in the DSA makes matters worse, and the black-letter nature of the act does not seem to be capable of preventing this dynamic from unfolding.[97]

Conclusion

The first generation of internet regulations, a model that provided absolute immunity for internet companies as its bedrock, led us to a crisis. A new model, with the DSA as its most salient example, is being developed. The bar is set higher for VLOPs and VLOSEs. While they retain their conditioned immunity, they have a new set of due diligence and transparency obligations that can make them responsible for noncompliance. The European Union is leading this new experiment to make businesses accountable for the consequences of their activities. This new regulation brought about the challenging and

'Human Rights in the Oil and Gas Industry: When Are Policies and Practices Enough to Prevent Abuse?' (2022) 61 Business & Society 1512 <https://doi.org/10.1177/00076503211017435> accessed 11 September 2024.

95 European Commission Digital Services Act (n 4), article 41.1.

96 Daphne Keller, 'The Rise of the Compliant Speech Platform' (*Lawfare*, 16 October 2024) <https://www.lawfaremedia.org/article/the-rise-of-the-compliant-speech-platform> accessed 5 November 2024.

97 Caroline Omari Lichuma, 'Mandatory Human Rights Due Diligence (mHRDD) Laws Caught Between Rituals and Ritualism: The Forms and Limits of Business Authority in the Global Governance of Business and Human Rights' [2024] Business and Human Rights Journal 1 <https://www.cambridge.org/core/journals/business-and-human-rights-journal/article/mandatory-human-rights-due-diligence-mhrdd-laws-caught-between-rituals-and-ritualism-the-forms-and-limits-of-business-authority-in-the-global-governance-of-business-and-human-rights/E8578EFE441CA76E61E461B0F2045A6D#fn3> accessed 5 November 2024.

much-needed debate over what tech company accountability could look like and how it could be enforced. It also brought the platform governance multi-stakeholder community out of gridlock and forced us to unpack and build consensus for the adoption of new terms of art and standards for this industry. The DSA has hence been effective already in many different ways.

During the implementation stage, however, the meaning of new terms needs to be interpreted and fleshed out. Best industry practices will be identified, and mandated risk assessment reports will ideally provide more nuanced information about content moderation and curation structures and practices within companies. The eyes of the world will be somewhat set in Europe to see if the model delivers what it promises.

This paper has tried to work through a number of open questions and identify inherent tensions in the risk-based approach adopted by the DSA that could hinder its effectiveness and may have broader impacts on the conception of the right to freedom of expression in Europe and beyond. While some of the challenges identified may be addressed during the implementation or through litigation, others probably cannot. The latter may nevertheless be important as the DSA gets –willingly or inadvertently– turned into a model for international soft law documents and comparative legislation.

As we analyzed, this approach is neither the logical corollary of applying the UNGPs to the ICT sector nor is it entirely consistent with international Human Rights standards of freedom of expression. It pushes rights out of the center stage and replaces them with risk analysis and the new governance techniques associated with the concept. Human rights become part of risk assessment processes, something that downplays their importance and efficacy while using their language and terminology to legitimize the new paradigm.

The enforcement stages of the DSA, however, offer opportunities to bring rights back to the center and make this new paradigm work. First, the European Commission should issue guidelines that demarcate the scope of companies' due diligence obligations under articles 34 and 35 of the DSA. Conversely, and even in the absence of those, when assessing the content-related risks stemming from their operations, VLOPs and VLOSEs should flesh out the open-ended terms and generic obligations by anchoring them to existing legal frameworks. For instance, the interpretation of "gender-based violence" and "the protection of minors", should be tied to the text of the CEDAW and the United Nations Convention on the Rights of the Child, and the full body of hard and soft law arising from the relevant treaties under the International and European systems of Human Rights protection, and the interpretation and application to concrete cases by the pertaining Tribunals and Committees. Furthermore, regardless of its absence in Article 34, companies should identify and assess, if appropriate, any risks stemming from

state action, including removal orders and more subtle "jawboning" attempts, in their reports under the DSA.

From a freedom of expression perspective, the indirect expansion of the speech to be governed by the state is incompatible with agreed international standards. Under the weight of systemic risks, content that is perfectly legal is subjected to the DSA's indirect speech governance mechanisms. The DSA allows the state to achieve, through the risk-based approach, what it could not legitimately do through a more direct and traditional form of regulation. While the risk-based approach to the regulation of VLOPs and VLOSEs responds to new potential sources of harm, carefully addressing volume, speed and permanence as potentially independent sources of legally redressable harm may be a first step in the right direction.[98] So far, volume, speed and permanence have been only indirectly addressed by States and current legal regimes do not contemplate them as independent sources of harm but rather as elements to determine remedies. Openly discussing these issues will allow for a more sincere and productive conversation within the platform governance community although it will bring about the need to reconsider, as Post suggests, some of the fundamentals of freedom of expression (like causality for instance).

Good democratic politics, in every single state of the Union, will have to dwell on the extent to which the DSA delivers its double promise of freedom and safety. Otherwise, we expect a healthy degree of pushback, driven—mainly—by concerned citizens and individuals. While premature, we can nevertheless imagine a future where the DSA incorporates obligations for states as well as companies, especially safeguards to prevent abuses from enforcement mechanisms and the specific identification of state actions that contribute to the "systemic risks" that companies are to address. Furthermore, restricting the DSA risks to those proposed within the human rights framework, thus, making it more like the UNGPs, would likely make the whole endeavor more narrow and viable.

98 Del Campo (n 38).