

Beatriz, Kira

Article

Regulatory intermediaries in content moderation

Internet Policy Review

Provided in Cooperation with:

Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

Suggested Citation: Beatriz, Kira (2025) : Regulatory intermediaries in content moderation, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 14, Iss. 1, pp. 1-26,
<https://doi.org/10.14763/2025.1.1824>

This Version is available at:

<https://hdl.handle.net/10419/315581>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/3.0/de/deed.en>



Volume 14 | Issue 1



RESEARCH
ARTICLE



OPEN
ACCESS



PEER
REVIEWED

Regulatory intermediaries in content moderation

Beatriz Kira *University of Sussex*

DOI: <https://doi.org/10.14763/2025.1.1824>

Published: 31 March 2025

Received: 10 April 2024 **Accepted:** 7 August 2024

Funding: This research was supported by Seed-Corn Funding provided by the School of Law, Politics and Sociology, University of Sussex.

Competing Interests: The author has declared that no competing interests exist that have influenced the text.

Licence: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License (Germany) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. <https://creativecommons.org/licenses/by/3.0/de/deed.en>
Copyright remains with the author(s).

Citation: Kira, B. (2025). Regulatory intermediaries in content moderation. *Internet Policy Review*, 14(1). <https://doi.org/10.14763/2025.1.1824>

Keywords: Content moderation, regulatory intermediaries, Digital Services Act, Brazil

Abstract: The modern digital public sphere requires effective content moderation systems that balance the interests of states, technology companies, and the public. This article examines how two pieces of legislation establish regulatory intermediaries in an attempt to strike this balance: the EU's Digital Services Act (DSA) with its out-of-court dispute settlement (ODS) bodies, and the proposed Brazilian Bill 2630/2020, which assigns supervisory responsibilities to CGI.br, a multi-stakeholder body. The analysis reveals that the design of regulatory intermediaries, including pre-existing structures like CGI.br, significantly impacts platform governance and user experiences. While the DSA's ODS model offers a framework for independent and user-friendly dispute resolution, the non-binding nature of its decisions limits its effectiveness. Similarly, concerns remain regarding CGI.br's limited financial and human resources, and its legal fragility, which could undermine its ability to fulfil the responsibilities the Bill assigns to it. By examining these models, the article offers insights for creating more effective participative online governance systems. It provides recommendations for the implementation of the DSA and informs ongoing legislative discussions in Brazil.

This paper is part of **Content moderation on digital platforms: beyond states and firms**, a special issue of *Internet Policy Review* guest-edited by Romain Badouard and Anne Bellon.

Introduction

In October 2024, Ireland's Digital Services Coordinator, the Coimisiún na Meán, announced the certification of Appeals Centre Europe (ACE), an out-of-court dispute settlement body for content moderation disputes (Coimisiún na Meán, 2024). The possibility for users to appeal platforms' decisions to an independent body was established by the Digital Services Act (DSA), which requires the largest platforms to offer users a mechanism to resolve disputes over content moderation through bodies certified by individual member states. The ACE was the first such body to be certified in Ireland, a jurisdiction hosting the European headquarters of many technology companies. The announcement also drew attention due to ACE's funding by the Oversight Board Trust—the same entity supporting Meta's Oversight Board, which reviews a small number of appeals regarding content on Facebook, Instagram, and Threads, escalated for review by users or by Meta itself.

While the ACE and the Oversight Board are distinct entities and operate independently,¹ they both function as regulatory intermediaries (Medzini & Levi-Faur, 2023), playing an important role in determining what content remains online and what is removed from platforms. This article examines the role of such intermediaries in content moderation systems. It examines how a new wave of regulations establishing hybrid forms of governance have recognised and incorporated non-state and non-corporate actors into platform decision-making, and discusses how these legally mandated, state-supervised structures constrain, shape, and guide content moderation, supporting forms of 'enhanced self-regulation' (Medzini, 2021, 2022).

The rise of content moderation intermediaries can be linked to the evolution of digital platforms and growing societal demands for mechanisms to improve their accountability. Historically, platform design and content oversight were primarily under the private control of the platforms themselves (Gillespie et al., 2020; Klonick, 2018). However, concerns about platform safety, the lack of transparency in content moderation, and the desire for more democratic online spaces have led

1. The press statement by the Oversight Board clarifies that "[t]he Appeals Centre is a separate legal entity to the Oversight Board" and that "[t]he two bodies will operate independently of one another and play distinct, but complementary, roles (Oversight Board, 2024)

to novel legislations aiming to establish government-led governance mechanisms. Notably, many of these emerging regulatory efforts recognise the need for a wider range of stakeholders to be involved in content moderation systems.

To examine the features and functions played by regulatory intermediaries this article compares two distinct institutional designs proposed by recent platform regulations. Firstly, it examines the model introduced with the Digital Services Act (DSA) implemented in Europe in 2022. This law governs online intermediary services, tailoring obligations to the player's role, size, and impact. Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) are, amongst other obligations, required to identify and manage 'systemic risks' and to engage with out-of-court dispute resolution bodies (DSA, Article 21).

The European case is then contrasted with the proposal under discussion in Brazil – the Internet Freedom, Accountability, and Transparency Bill (PL No. 2630/2020, colloquially known as the 'Fake News Bill'). This proposed regulation seeks to structure a public-managed 'system for digital integrity and transparency'. Amongst other provisions, it allocates responsibilities to the Brazilian Internet Steering Committee (CGI.br), a multi-stakeholder internet governance body, which would be in charge of developing studies, issuing opinions, and suggesting codes of conduct for platform compliance.²

By identifying both the European law and the Brazilian bill as attempts to legally recognise and legitimise the role of regulatory intermediaries in the specific context of content moderation, this article seeks to contribute to the understanding of the role that actors beyond states and platforms play in governing online spaces. To do so, the article is structured as follows. Section 1 provides an overview of existing scholarship on civil society organisation and multi-stakeholder participation in platform governance. The literature shows how platform-led efforts to foster participation channels and acknowledge diverse voices, particularly those of marginalised communities, have yielded limited results. Furthermore, the literature also highlights the shortcomings of decentralised or federated content moderation systems in effectively protecting users. This suggests that a form of state-led regulatory effort is necessary to increase platform accountability and transparency. Hybrid systems that subject platforms' private power to forms of state oversight could potentially address some of these shortcomings, making content moderation more responsive to the concerns of diverse stakeholders.

2. The analysis is based on the most recent publicly available version of the Bill, presented by the rapporteur of the special committee, Orlando Silva, on 27 April 2023, available at https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2265334 (access 10 April 2024).

Section 2 introduces two cases of hybrid regulation that create space for regulatory intermediaries. It draws on the framework developed by Medzini and Levi-Faur to analyse provisions in the DSA and the Brazilian Bill 2630/2020. It argues that both legislations seek to establish ‘enhanced systems of self-regulation’ (Medzini, 2021, 2022) through the formal establishment and empowerment of ‘regulatory intermediaries’ (Medzini & Levi-Faur, 2023). Section 3 develops a comparative analysis of these legislations, first examining the delegation mechanisms established by these legislations that ensure a level of autonomy from the platform, and then examining the mechanisms created that seek to build credibility to the regulatory intermediaries, that is, to ensure they are up to their assigned regulatory task.

The article concludes by highlighting the importance of law and legal frameworks in structuring regulatory intermediaries for content moderation. It acknowledges that the effectiveness of these frameworks largely depends on their design and argues that while such frameworks can help to reduce the power imbalance faced by third parties attempting to regulate platform harms, they also introduce complexity, potentially hindering the implementation of the regulations. This emphasises the need for ongoing research and monitoring to examine the effectiveness of these regulations as they are implemented.

Section 1. From private to hybrid systems of governance

Early internet laws, in particular the intermediary liability laws adopted in the US and Europe, have largely exempted platforms from responsibility for user-generated content while granting them broad leeway for self-regulation. Platforms benefited from broad or conditional immunities from legal liability (Frosio, 2022). At the same time, concerns about user retention and brand safety have incentivised the largest social media companies to create ‘sanitised’ online spaces (Griffin, 2022) and further contributed to the development of sprawling and complex private governance systems on these platforms (Douek, 2022; Klonick, 2018, 2023).

As the scale of content moderation expanded, so did the need for more specific and granular policies and means of enforcing them (Klonick, 2018, pp. 1633–1634). These rules often went beyond prohibiting illegal or harmful content (Gillespie, 2018). For instance, many major social media platforms ban nudity and all forms of sexual content – likely suppressing large amounts of valuable and harmless content and disproportionately affecting marginalised users (Griffin, 2022).

While some argue that describing content moderation systems as ‘rough online analog of offline judicial adjudication of speech rights’ (Douek, 2022, p. 528) is a flawed and ‘stylised picture’ of content moderation, there are undeniable parallels between these systems and state-led decision-making and governance. Platforms do apply ‘legislative-style rules drafted by platform policymakers to individual cases’ (Douek, 2022, p. 529), exercising a form of adjudication, and these decisions constitute a crucial element of their systems of moderation. However, these complex, *quasi*-legal online content management systems are not administered by the state but by private companies, constituting a ‘private self-regulatory system to govern online speech’ (Klonick, 2018, p. 1631).

Over time, partly driven by scholars urging platforms to ‘take on an expanded sense of responsibility’ and ‘share the tools to govern collectively’ (Gillespie, 2018, p. 212), platforms have built structures and mechanisms aiming to make their content moderation systems more receptive to the concerns and priorities of other stakeholders. This involves a series of mechanisms, ranging from allowing users or institutions to ‘flag’ content on their platforms, partnering with fact-checking organisations, to more institutionalised formats such as multi-stakeholder councils. Caplan argues this strategy – developed partly in response to concerns about platforms’ unilateral control – serves a dual purpose: first, it gestures towards inclusivity and participation in content policymaking, and second, it aims to distribute the responsibility for policymaking and address resource and functional gaps inherent in operating at scale (Caplan, 2023). Importantly, this ‘networked governance’ approach transcends a dichotomous view of platform governance that ‘have tended to emphasize unilateral or unidirectional forms of control and influence’ (Caplan, 2023, p. 3454). By incorporating other actors, it moves beyond a focus solely on the role of private or public actors.

However, the literature raises concerns about the structure and control of these participation channels, and who ultimately benefits from them. While identifying various avenues for civil society involvement in content moderation, Bietti highlights a key weakness: such involvement is often framed as a form of platform self-regulation (Bietti, 2023). This means participation is defined and controlled by the platforms themselves, raising concerns about the independence of the participating organisations (Bietti, 2023, pp. 42–43). As Bietti argues, ‘individual privileges to occupy privately-controlled spaces are granted through emulations of the rule of law and otherwise subject to platforms’ discretion. Private power is left to operate freely in the background’ (Bietti, 2023, p. 45).

Similarly, Griffin critiques multistakeholderism, which empowers other actors to

pressure platforms but often emphasises voluntary corporate social responsibility commitments, thus minimising direct state regulation (Griffin, 2023, p. 48). Griffin argues that platforms operate within an unequal economic system and that civil society and multi-stakeholder responses shaped by platforms ultimately exacerbate these inequalities. Griffin points out that governance models aiming to 'make private institutions more permeable to influence from civil society and other market actors inevitably produce unequal outcomes, given enormous disparities in resources and influence among civil society groups, consumers and other stakeholders' (Griffin, 2023, p. 71). Therefore, these multistakeholderism efforts would be 'unlikely to produce outcomes that genuinely challenge these companies' economic interests' (Griffin, 2023, p. 74).

Empirical studies support this scepticism. Analyses of the Airbnb and Facebook civil rights audits, conducted by Baik and Sridharan (2023) using interviews and textual analysis, reveal the limited effectiveness of integrating civil rights frameworks within platform structures. These studies highlight the challenges faced by such audits due to the inherent power imbalances. The authors argue that these challenges contribute to 'the externalization of responsibility by platforms' and 'the co-option and politicization of civil rights', ultimately failing to advance the protection of rights for marginalised communities (Baik & Sridharan, 2023).

There are, however, more optimistic perspectives on the involvement of third parties in content moderation systems. For example, Heldt and Dreyer argue that establishing independent bodies to make content-related decisions would increase public accountability, provided these bodies are neutral, independent decision-makers, and transparent in their decision-making processes. They focus on the role of these third parties in content moderation and argue that 'a competent third party independent from both governmental and corporate interests that is equipped with powers to decide on content moderation matters could be the way forward' (Heldt & Dreyer, 2021).

Examples of such third parties could include Twitter's Trust and Safety Council (which was dissolved by Musk in December 2022) and TikTok's Content and Safety Advisory Councils. However, the most developed example of an institutionalised, platform-led multi-stakeholder organisation is arguably Meta's Oversight Board. Established in October 2020 by Meta (then Facebook), this *quasi*-adjudicatory body reviews selected content moderation decisions (Klonick, 2020). Established as a private entity and managed by a trust, the Board operates under a set of privately established rules. Its Charter, likened to a 'Constitution', and its Bylaws, serving as the equivalent of procedural court rules, govern the Board's operations and its cor-

porate relationship with Meta.³

The Oversight Board model, with its power to overturn content moderation decisions, its substantial resources, and its high-profile members, has attracted significant attention in the literature. Critics have noted that the Board selects only a tiny fraction of the immense number of content moderation decisions appealed for review and can only make binary decisions on whether to remove or keep up content, thereby missing the opportunity to recommend more proportional measures, such as demotion (Howard & Kira, 2024). Additionally, the Board's effectiveness, mandate, and operations have been questioned, focusing specifically on the Board's financial and practical independence (Douek, 2024; Klonick, 2020). Notably, the Board is funded by a trust established by Meta. While all 20 founding members, from diverse geographical backgrounds and disciplines, were also appointed by Meta (Bietti, 2023, p. 43), the Board has since transitioned to a model where it selects all future members itself.⁴ More broadly, concerns have been raised regarding the Board's application of international human rights law and its lack of democratic legitimacy due to the limited scope for effective civil society participation in its decision-making process (Dvoskin, 2023).

Recognising the limitations of platforms controlling participation terms and oversight for other actors, new governance models have emerged, aiming for a more effective decentralisation of platform power. The literature has subsequently examined the design, motivations, and effectiveness of alternatives to corporate-controlled social media. Discussions have centred primarily on decentralised or multi-level platforms, often focusing on innovative models that seek to offer users a plurality of content moderation choices. For example, Jhaver and colleagues (2023) catalogue multi-level platform designs to highlight the variety of available governance structures. They show that there are significant 'differences in governance design, or the ways in which governance is intended to be carried out according to platform creators, designers, and implementers' (Jhaver et al., 2023). They contrast centralised, corporate-driven models like X (formerly Twitter) with alternatives such as Mastodon and Bluesky. These federated models, which gained traction particularly after Elon Musk's acquisition of Twitter, are designed to offer more decentralised governance, with user-led moderation and curation tools (Jhaver et al.,

3. The original documents were drafted exclusively by Facebook's employees, with the creation of the Board tasked to Facebook's Governance and Initiatives Team. A detailed account of the creation of the Board is provided in (Klonick, 2020).

4. See 'Updates on Oversight Board membership', available at <https://www.oversightboard.com/news/771690787717546-updates-on-oversight-board-membership/> (posted on April 2023, accessed 7 July 2023)

2023).

While offering a potential alternative to corporate-controlled platforms, decentralised governance presents its own set of challenges for content moderation. Research suggests these platforms struggle to tackle issues like misogyny, networked harassment, and image-based abuse (Marwick & Caplan, 2018). Further, compared to centralised platforms federated models reportedly face significant difficulties in establishing robust and scalable moderation systems, particularly for persistent threats like spam and coordinated attacks (Roth & Lai, 2024). Key challenges include 'gaps in platform moderation capabilities' and lack of sustainable funding to develop trust and safety work (Roth & Lai, 2024). This highlights the need for innovative solutions that address both the limitations of private systems and the challenges inherent in decentralised approaches.

Legally requiring platforms to engage with third-party organisations in content moderation could offer an effective solution to concerns about participation without compromising on safety. So far, platform-led forms of participation in centralised platforms have largely taken place as part of what Medzini and Levi-Faur term 'thin self-regulation', where platforms retain control over the mechanisms of participation, transparency, and accountability of third parties (Medzini & Levi-Faur, 2023). These self-regulatory models, as Bietti highlights, often lack 'a thick conception of constraint on private power and an emphasis on the effects of concentrated private commercial power on democracy' (Bietti, 2023, p. 45). Similarly, a central aspect of Caplan's critique of existing platform governance networks is that they lack legal structures and are not established by formal laws (Caplan, 2023, p. 3465).

This suggests that involving a form of state-led legislation requiring public, democratic oversight of participation mechanisms could be a viable approach. Indeed, it seems plausible that legally structuring channels for the inclusion of civil society organisations and other actors in content moderation decision-making processes can make content moderation more inclusive of interests beyond those of the platforms themselves, avoiding the shortcomings of private-led initiatives. While a degree of centralisation would be necessary to ensure proper development and resourcing, these mechanisms would also be subject to state oversight to prevent capture and guarantee accountability. This hybrid model of content moderation governance is precisely what emerging platform regulation laws have sought to achieve.

In summary, as social media platforms become larger and more complex, new and

more challenging problems arise regarding what content should be allowed and how much exposure it should receive. Given the sheer volume of content shared on these platforms, content moderation systems designed to make such decisions have evolved to become increasingly complex, relying heavily on automated mechanisms. However, as the challenges of content management grow, the limitations of private-led forms of regulation become more apparent. Public pressure has also mounted for these decisions to be subject to more democratic accountability, rather than solely resting with platforms. This has led to experiments in decentralising decision-making and recognising the participation of not only platforms and regulators but also third parties in content moderation.

Section 2. Regulatory intermediaries in platform regulation

In recent years, a new wave of platform regulation has emerged, aiming to hold platforms more accountable. These regulations were developed partially in response to mounting evidence on the prevalence of illegal and harmful online content (Buiten, 2021; Woods, 2019), along with reports from whistleblowers and former platform employees that have exposed the limitations of a pure, or ‘thin’ self-regulatory regimes (Medzini & Levi-Faur, 2023).

While some jurisdictions have opted for direct government control over content (e.g. Turkey, Russia, China), or even platform bans (e.g. US ban on TikTok), the most promising models emerging are hybrid forms, known as co-regulation (Keller, 2022). This approach subjects platforms’ private governance systems to some form of government oversight, combining different types of rules (soft and hard law). Some of these regulations, exemplified by the EU Digital Services Act and the Brazilian Bill 2630/2020, push the boundaries of ‘hybrid’ further by introducing a novel element: regulatory intermediaries. These are non-governmental, non-platform actors introduced into content moderation frameworks, with roles recognised and assigned by law.

Given the novelty of this model, it is important to understand the implications of these content moderation regulatory intermediaries. How do they fit into the wider regulatory approach these new legislations are seeking to establish? Will the fact that they are established in law, rather than voluntary, mitigate the issues raised in the literature? How can state-led legislation effectively structure these channels of participation in decision-making? Through a legal analysis of the approved Article 21 of the EU DSA and the proposed Brazilian Bill 2630/2020, the following subsections explore the legal choices made by lawmakers to decentralise or delegate

platform power and how the regulatory intermediaries are designed and structured.

2.1 Theoretical framework

The concept of regulatory intermediation proposed by Abbott, Levi-Faur, and Snidal views regulation as a three-party system involving a regulator (R), the targets of regulation (T), and intermediaries (I) (Abbott et al., 2017b). This triangular model of regulation has been extensively explored in various domains, including food safety (Lytton, 2017), financial markets (Kruck, 2017) and marine fishery certification (Auld & Renckens, 2017). While the literature shows the RIT model's versatility and highlights the diverse roles, formats, and functions of intermediaries in different regulatory contexts, it also identifies common concerns. Notably, intermediaries often pursue their own private interests, making the examination of these interests – along with the regulatory design features that influence them – crucial for understanding their participation in regulation (Abbott et al., 2017a).

With a focus on regulatory design and digital regulation, Medzini and Levi-Faur (2023) examine how two cyberspace polycentric regimes have established 'enhanced self-regulation' by delegating responsibilities to regulatory intermediaries: the delegation within the EU General Data Protection Regulation (GDPR) and the private delegation seen in Facebook's content moderation (Medzini & Levi-Faur, 2023). In these regimes, intermediaries 'facilitate, manage, and improve the credibility of self-regulation in complex polycentric governance regimes' (Medzini & Levi-Faur, 2023, p. 323). Relevantly, they show that in the digital context regulatory intermediation 'symbolizes a transition away from two-party modes of regulation', moving the debate towards 'polycentric relationships that include various groups of actors' (Medzini & Levi-Faur, 2023, p. 326).

As such, the findings of Medzini and Levi-Faur on the strengths and weaknesses of enhanced regulation models are intended as a blueprint for establishing hybrid forms of enhanced self-regulation, particularly regarding the way these arrangements are designed. Crucially, they identify two key shortcomings to avoid in designing enhanced self-regulation: a lack of autonomy from appointing actors and the potential failure to achieve long-term goals (Medzini & Levi-Faur, 2023).

Drawing on their findings, this article examines the design and empowering role of 'regulatory intermediaries' in the EU DSA and in the Brazilian Bill 2630/2020. Specifically, the legal analysis focuses on the legal choices concerning two key dimensions of regulatory intermediaries that correspond to the shortcomings identi-

fied by Medzini and Levi-Faur: independence and credibility, and the need to balance these two aspects (Medzini & Levi-Faur, 2023, p. 328). By applying this framework to two content moderation regimes in early stages of development, this analysis aims to inform lawmakers and policymakers and help them refine their approaches to improve the effectiveness of their regulatory intermediaries.

2.2 Case studies

The Digital Services Act is the European Union's flagship legislation governing online intermediary services. Approved in 2022, it has been through a staged implementation process to allow regulated entities time to prepare, with the bulk of the obligations coming into force on 17 February 2024.

The DSA establishes a differentiated approach to platform regulation, so that the specific regulatory requirements adapt to the type, size, and nature of the intermediary service concerned. The regulation sets out core obligations for all providers of intermediary services, with additional obligations for hosting services, online platforms in general, and further, more stringent rules for very large online platforms (VLOPs) and very large online search engines (VLOSEs). The EU Commission is responsible for enforcing the law alongside member states, which should set up national Digital Services Coordinators to supervise compliance for platforms established within their territory. The Commission, however, takes the lead role in monitoring and enforcing the stricter obligations applicable to VLOPs and VLOSEs.

A crucial element of the DSA is its strong emphasis on fair procedures and individual rights. This is reflected in platform obligations, such as publishing clear and accessible content policies, applying these policies consistently with due consideration of users' fundamental rights, publishing regular public transparency reports, and operating internal appeals processes (Griffin, 2023, pp. 57–58). In this sense, a key innovation within the DSA's procedural requirements is the mandate for platforms to provide users with access to independent out-of-court dispute settlement (ODS) bodies.

Article 21 of the DSA specifically grants all recipients of online platform services, including both individuals and entities, the right to access out-of-court dispute settlement bodies. This right applies in relation to disputes concerning platform decisions – including decisions related to content removal, de-amplification, suspensions, and content monetisation – and in relation to other complaints that have not been resolved through the platform's internal complaint system. The article states that:

Recipients of the service, including individuals or entities that have submitted notices, addressed by the decisions referred to in Article 20(1) shall be entitled to select any out-of-court dispute settlement body that has been certified in accordance with paragraph 3 of this Article in order to resolve disputes relating to those decisions, including complaints that have not been resolved by means of the internal complaint-handling system referred to in that Article (DSA, Article 21).

Article 21 also mandates that online platforms must ensure information about accessing ODS bodies is readily available on their user interfaces, presented in a clear and user-friendly manner. This applies to all online platforms, encompassing those that bring together sellers and consumers, such as online marketplaces, app stores, collaborative economy platforms, and social media platforms.

Brazil's proposed platform regulation bill, Bill 2630/2020, also establishes a regulatory intermediary with a distinct design and function. Introduced in 2020 amid concerns about pandemic-related misinformation on platforms (Machado & Aguiar, 2023), Bill 2630 has faced criticism for potentially stifling free expression while also being deemed insufficient for user protection and misinformation control (Keller & Dos Santos, 2024). Notably, the challenges associated with content moderation and countering the spread of harmful content are exacerbated in Global South jurisdictions like Brazil, given inequalities in the distribution of content moderation resources (Udupa et al., 2023, p. 11) and inaccuracies in fact-checking in languages other than English (Vinhas & Bastos, 2023).

Particularly over the past few years, platform liability and content moderation have become increasingly pressing issues in Brazil. The storming of the Brazilian capital by far-right supporters of former President Jair Bolsonaro on 8 January 2023, which was partially orchestrated using social media and messaging apps, prompted the new Brazilian government to actively pursue measures increasing the responsibility of intermediaries in moderating online content (Hartmann, 2023). The legal disputes between the Brazilian Supreme Court and Elon Musk, and the lack of compliance with court orders requiring the suspension of user accounts on X – which led to the temporary suspension of the platform in Brazil in September 2024 – are further examples of attempts to subject platforms to the rule of law in Brazil (Jost & Cruz, 2024).

Similarly to the DSA, the Brazilian Bill adopts a risk-based approach. According to the proposal, platforms must identify, analyse, and diligently assess systemic risks

arising from their services' design, operation, and related systems, including algorithms. Based on this assessment, they are required to implement reasonable, proportionate, and effective mitigation measures. Additionally, providers must develop a code of conduct with qualitative and quantitative indicators to ensure compliance with the proposed law.

The bill assigns key oversight responsibilities to CGI.br, a multi-stakeholder body that is not a government entity (although it includes government representatives). Among other responsibilities related to developing studies and debates, the bill seeks to task CGI.br with duties related to supervising the proposed law's hybrid system of governance. Article 51 of the bill grants CGI.br specific powers, including issuing guidelines for social media providers, search engines, and instant messaging platforms to develop codes of conduct, and validating these codes after they have been drafted by the platforms.

Additionally, CGI.br will be responsible for publishing a list of providers that fall within the scope of the regulation, based on the criteria established in the proposed legislation. It will also issue preliminary recommendations before the initiation of any administrative proceedings in cases where the information contained in transparency reports is insufficient or where an independent audit finds the reports to be unsatisfactory. Critically, CGI.br will issue guidelines and requirements for the analysis of systemic risks and will analyse the providers' systemic risk assessment reports.

TABLE 1: Regulatory intermediaries in the EU DSA and the Brazilian Bill No. 2630/2020

	EU DIGITAL SERVICES ACT	BRAZILIAN BILL NO. 2630/2020
REGULATORY INTERMEDIARY	Certified out-of-court dispute settlement (ODS) bodies	Brazilian Internet Steering Committee (CGI.br)
REGULATORY INTERMEDIATION ROLE	Rule enforcement	Rule-making (guidelines and validation of codes of conduct), rule monitoring (analysis of systemic risks reports), and rule enforcement (initiation of administrative procedures)
LEGAL BASIS	EU Digital Services Act (article 21)	Bill No. 2630/2020 (proposed article 51)

	EU DIGITAL SERVICES ACT	BRAZILIAN BILL NO. 2630/2020
STATUS	Adopted in November 2022, in force	Introduced in 2020, at the time of writing was under examination by Brazilian Congress

Both the EU’s Digital Services Act and Brazil’s Bill 2630/2020 share a similarity: they require large social media platforms to involve non-governmental actors in key parts of their content moderation systems. In both cases, the legal framework seeks to establish a regime of ‘enhanced self-regulation’ through formal intermediaries, established and structured by binding legal provisions. These approaches, however, diverge in terms of the level of governance they impact and how they structure third-party participation. This variation can be valuable for comparative analysis, allowing the assessment of the strengths and weaknesses of different models in promoting independent yet credible models of enhanced self-regulation.

Section 3. Comparative analysis of regulatory intermediaries in EU DSA and Brazilian bill

The comparative analysis delves into two dimensions: independence and credibility, which are identified by the literature as being central for the design of effective regulatory intermediaries. Examining independence, it assesses how the legislation delegates responsibilities and discretion to intermediary actors. Here, the focus is on how the legislation establishes a framework to ensure these intermediaries ‘do not drift away or get captured’ by the platforms they oversee (Medzini & Levi-Faur, 2023, p. 328). The second is credibility, focusing on ‘the manner in which policymakers balance between permitting private actors to act as independent intermediaries and establishing an institutional regulatory design that ensures the long-term credibility of the intermediaries’ (Medzini & Levi-Faur, 202, p. 329). This analysis entails examining the legal constraints, requirements, and criteria for certification of the regulatory intermediaries, exploring how these are established, and the methods used to ensure the intermediaries fulfil them.

3.1. The EU Digital Services Act and out-of-court dispute settlement

The DSA establishes clear requirements for ODS bodies to act as regulatory inter-

mediaries. These bodies must be impartial and independent, including financial independence from both online platform providers and service recipients, encompassing individuals or entities submitting notices (Article 21(3)(a)). Additionally, the DSA requires members to be remunerated in a way that is not linked to the outcome of the procedure (Article 21(3)(c)).

Furthermore, certified bodies must handle complaints relevant to all platforms, not solely those associated with the body itself (Wimmers, 2021). This contrasts with ‘thin’ forms of self-regulation, discussed in section 1, where each platform established and funded its own council or oversight institution, which lacked autonomy and risked interference. The ODS model proposed by the DSA makes ‘capture’ more difficult by disassociating the review body from a specific platform. Additionally, this model establishes market-based incentives that are likely to lead to better services from ODS bodies. By allowing users seeking to use ODS to select any certified body within their jurisdiction, the legislation creates a market for independent review bodies, fostering competition. While this would be a new market and could also lead, at least in the short term, to competition for resources and experts, the growing relevance of trust and safety could attract more investment and talent in the medium to long term. There is also already evidence of initiatives that seek to foster a collaborative culture between certified ODS, such as the network of independent out-of-court dispute settlement bodies announced by ACE and User Rights.⁵

The independence of the regulatory intermediary also hinges on its ability to enforce its decisions without resorting to government intervention or relying solely on the goodwill of regulated entities. In this sense, a crucial limitation of the model introduced by Article 21 is that the outcome of the ODS procedure is not a binding settlement of the dispute between the parties. While the DSA originally proposed that platforms would be bound by the decisions of certified bodies (Ortolani, 2022), the final version of the law merely requires platforms to provide users with clear and user-friendly information about the possibility of using ODS mechanisms to resolve disputes, and require both parties to ‘engage, in good faith, with the selected certified out-of-court dispute settlement body with a view to resolving the dispute’. However, there is no requirement to commit to the outcome of the proceedings. While this does not render the entire regime a ‘thin’ form of self-regulation, it certainly weakens its effectiveness.

5. See LinkedIn post by the Appeals Centre Europe from 20 October 2024: https://www.linkedin.com/posts/appealscentre_in-the-coming-weeks-we-will-be-announcing-activity-7252301629809393667--DXM (accessed on 3 November 2024).

However, even if the outcome is not binding, the interaction between ODS and platforms can ultimately lead to more efficient content moderation systems. As Husovec argues, ‘the DSA does not take away all the content moderation discretion from platforms’ so that ‘if providers do not like how out-of-court bodies read their rules, they can change them and make them clearer’ (Husovec, 2023, p. 118). This suggests that the activities of regulatory intermediaries can ultimately foster an iterative process of revising content moderation practices.

It is important to notice that while ODS bodies are required to be independent from platforms, they don’t necessarily have to be independent from the state. Article 21(6) allows member states to establish or support ODS bodies, as long as they do not interfere with the ability of Digital Services Coordinators to certify these bodies.

In terms of credibility, Article 21 of the DSA establishes a very specific framework for regulatory intermediaries, outlining detailed requirements for certification as ODS bodies. Beyond the requirement for impartiality and independence, ODS bodies must meet specific quality-related criteria. For example, the law mandates that they possess the ‘necessary expertise in relation to the issues arising in one or more particular areas of illegal content, or in relation to the application and enforcement of terms and conditions of one or more types of online platform, allowing the body to contribute effectively to the settlement of a dispute’.

Furthermore, the DSA establishes several key characteristics that out-of-court dispute settlement must embody. It must be easily accessible through electronic communication technology, allowing users to initiate disputes and submit supporting documents online. Additionally, the process should be swift, efficient, and cost-effective, conducted in at least one of the official languages of the European Union institutions. Finally, the ODS body must operate under clear and fair rules of procedure that are easily accessible to the public and comply with applicable law.

To ensure they meet these criteria, ODS bodies must undergo a certification process conducted by national Digital Service Coordinators (Casarosa, 2023). This process aims to guarantee that the bodies function as genuine third-party entities with the necessary expertise to adjudicate relevant disputes. However, ODS bodies can only be certified in the nation where they are established, raising questions about the applicability of these certifications in other countries (Casarosa, 2023, p. 45).

This *national* certification process has also sparked concerns about potential incon-

sistencies in the quality of ODS bodies. While local accreditation might ensure bodies are prepared to interpret and adjudicate local law (which is important as il-legal content is decided based on the law of each member state), it could lead to variations in decision-making procedures, expertise, and ultimately, the quality of dispute resolution. As Wimmers argues, the lack of ‘standards or criteria for the complex factual and legal determinations and balancing of rights in the area of online speech’ could lead to significant variation in outcomes (Wimmers, 2021, p. 382). It could be possible, for example, that similar cases involving the same platform are decided very differently by ODS based on different member states.

Another concern is more generally the suitability of ODS for adjudicating rights-based disputes, particularly those involving freedom of expression. ODS mechanisms typically prioritise consensual solutions based on interests rather than legal positions or asserted rights. This raises worries that models designed for commercial disputes might be ill-equipped to handle fundamental rights issues (Wimmers, 2021, p. 382).

The current landscape suggests that new entities are being established to meet the DSA’s requirements, including one building on the expertise of the Meta’s Oversight Board, which would not itself qualify under the DSA’s criteria. The Board’s structure directly contradicts the requirements for independence and transparency outlined in Article 21. It is funded and established by Meta, operates exclusively on its platforms, and has the discretion to select which cases it hears (focusing on a very small proportion of appeals or referrals).

Accreditation procedures are being set by European Digital Services Coordinators that seek, amongst other criteria, to balance the independence and credibility of ODS bodies. For example, the guidance issued by the Coimisiún na Meán, the Irish regulator, to help applicant ODS bodies provides details regarding the information that bodies should provide to evidence impartiality and independence, financial independence, the expertise of the ODS body and its dispute resolution personnel, and how this expertise will be maintained and enhanced over the period of certification. Such aspects are crucial to ensure that an ODS is able to function as an effective regulatory intermediary and contribute to enhanced models of self-governance.

3.2. The Brazilian Bill 2630/2020 and the Brazilian Internet Steering Committee (CGI.br)

The Brazilian bill leverages an existing body, the Brazilian Internet Steering Com-

mittee (CGI.br), granting it regulatory intermediary status. Founded in 1995, CGI.br is a multi-stakeholder organisation. Restructured by a presidential decree (Decree No. 4829/2003) in 2003, CGI.br currently has 21 members. These include nine government representatives, four from the business sector, four from civil society organisations (the third sector), three from the scientific and technological community, and one with recognised expertise in internet affairs. Since July 2004, non-governmental representatives have been democratically elected by their peers for a three-year term, with the possibility of re-election.

CGI.br's long history of supervising internet governance in Brazil indicates its expertise and lends it credibility. Even before Bill 2630/2020 was passed, CGI.br already held some regulatory responsibilities, albeit less developed than what the bill proposes. For instance, Decree No. 4829/2003 formally assigns CGI.br the responsibility for establishing strategic guidelines for internet use and development, managing Brazil's top-level domain names ('.br') and IP addresses, proposing internet-related research and development programmes, and coordinating actions related to proposing regulations for internet activities, among others.

Beyond its internet governance responsibilities, CGI.br has some experience in platform regulation specifically. In 2009, it developed principles for internet governance and use in Brazil (Resolution CGI.br 2009/003/P) which significantly influenced the drafting of the landmark piece of legislation known as Brazil's Internet Civil Rights Framework (*Marco Civil da Internet* - MCI). Approved in 2014, following a comprehensive public participation process, the MCI (Law No. 12965/2014) enshrines the protection of users' fundamental rights online, particularly freedom of expression and privacy. It also assigns further responsibilities to CGI.br. For instance, the MCI states that the 'promotion of rationalised internet management, expansion, and use' will involve the participation of CGI.br, and that the committee will be consulted on the application of net neutrality rules, specifically concerning the technical requirements for 'discrimination or degradation of traffic' (MCI, Article 9).

This experience in online regulation positions CGI.br favourably to undertake the intermediary regulatory activities proposed by Bill 2630/2020. Indeed, its expertise in this area was highlighted during public hearings, with civil society organisations advocating for CGI.br to be designated as the supervisory body more broadly. The report from the committee that examined the legislative proposal cited contributions from public hearing participants, stating that CGI.br 'could assume the attributions foreseen in the bill, leveraging the expertise already developed by CGI.br in regulating and supervising the functioning of the internet in Brazil' (Spe-

cial Committee on Bill No. 2630/2020, 2023, p. 26). The report further noted that participants argued that ‘CGI.br would be the ideal forum to carry out these attributions, given its nearly three decades of experience in multi-stakeholder internet governance’ (Special Committee on Bill No. 2630/2020, 2023, p. 26).

In terms of credibility, CGI.br would not have to face the same challenge of ODS in navigating a fragmented legal landscape. While in Europe regulatory intermediaries will have to operate under both the laws of individual member states where they are certified and European-level rules, in Brazil the relevant laws apply nationally. Even though Brazil is structured as a federation, laws related to the regulation of telecommunications and media regulation cannot be made by states and municipalities – these subjects can only be regulated at the national level. This is the case of the Marco Civil da Internet and will also be true if bill 2630/2020 (or a version of it) is adopted in the future. Of course, the interpretation and application of laws by local courts can vary across states, but the higher courts have nationwide jurisdiction.

While fragmentation is not a key concern in the Brazilian context, the fact it is a Global South jurisdiction that faces greater budgetary constraints than the EU gives rise to a different set of challenges. Crucially, there are concerns regarding CGI.br’s capacity to effectively function as a regulatory intermediary in the terms proposed by Bill 2630/2020.⁶ Public hearing participants also highlighted that ‘the attributions given by the proposal to the steering committee are not alien to CGI.br’s current activities’. Nevertheless, they emphasised concerns from committee members regarding the additional responsibilities, especially considering its limited structure (Special Committee on Bill No. 2630/2020, 2023, p. 30). This lack of structure is linked to questions about CGI.br’s independence. As a multi-stakeholder organisation, it operates independently of the Brazilian government and platforms, serving a dual role: as a ‘forum of debate and deliberation between stakeholders’ and ‘as one stakeholder among many within the Brazilian Internet ecosystem’ (Anastácio, 2018).

CGI.br’s high level of institutional independence, however, brings challenges in terms of its financial sustainability. The Committee’s non-governmental status limits its capacity and funding sources. Members provide their time and expertise voluntarily, alongside their regular jobs. None of them are salaried employees of CGI.br. These are significant shortcomings in CGI.br’s design, predating Bill 2630/

6. For a broader discussion about capacity constraints in the context of digital regulation in Brazil, see Kira (2024).

2020, that impact its ability to function effectively as a regulatory intermediary under the terms of the Bill.

Furthermore, CGI.br's legal status is delicate. Monteiro Neto argues that the committee 'has a very fragile legal status as it can be modified or extinguished by the enactment of another normative instrument without the need for discussion in parliament' (Monteiro Neto, 2018, p. 56). The decree establishing its composition and responsibilities is short and vague, with many of its rules and functions based on custom. Also, because it was established by decree, its composition and functions can also be changed unilaterally by presidential decree, further compromising its independence.

These issues raise questions about CGI.br's suitability to perform the significant regulatory intermediary role proposed in the Brazilian bill. The sheer volume of non-trivial responsibilities outlined in the bill could prove difficult to implement effectively given CGI.br's existing limitations – it would require a significant review and an increase in resources and power for this model to be effective.

TABLE 2: A comparison between regulatory intermediaries in the EU DSA and in the Brazilian Bill

	EU DSA	BRAZILIAN BILL
MODE OF ENHANCED SELF-REGULATION	HIERARCHICAL AND MARKET-BASED	HIERARCHICAL AND NETWORK-BASED
REGIME FORMALITY	FORMAL AND LEGALISED VIA DIGITAL SERVICES ACT + CERTIFICATION PROCESS SET BY DIGITAL SERVICES COORDINATORS	FORMAL AND LEGALISED VIA FUTURE INTERNET FREEDOM, ACCOUNTABILITY, AND TRANSPARENCY LAW
WHO SETS THE CRITERIA/RULES?	DIGITAL SERVICES COORDINATORS	LAWMAKERS (WITH REGARDS TO RESPONSIBILITIES) AND EXECUTIVE POWER (WITH REGARDS TO COMPOSITION AND GENERAL ATTRIBUTIONS)
CAN REGULATORS DIRECTLY REGULATE INTERMEDIARIES VIS-À-VIS THE CRITERIA?	YES	YES
CAN REGULATORS DIRECTLY REGULATE REGULATED ORGANISATIONS VIS-À-VIS THE CRITERIA?	No	No

Conclusion

In light of the emergence of hybrid systems of governance, there is a need for further examination of the legal nature, structure, and function of regulatory intermediaries. These intermediaries introduce non-platforms non-state actors within content moderation systems, with implications for online speech governance. This article investigated how law structures the roles of these intermediaries, focusing on out-of-court dispute settlement mechanisms under the EU DSA and the role assigned to the multi-stakeholder body CGI.br proposed by the Brazilian bill. It compared the implications of these legislative choices in filling the democratic gap of platform governance and how they fit within the wider co-regulatory frameworks proposed by each legislation.

It found that the EU DSA's out-of-court dispute settlement model is formal, with legal criteria for certification and a hierarchical structure subject to approval by national Digital Services Coordinators. It also benefits from market-based incentives, as the possibility of multiple ODS bodies serving multiple platforms creates competition. Overall, the DSA's ODS model offers a framework with strong potential for fostering independent and user-friendly online dispute resolution. However, limitations around the enforceability of decisions, the suitability for rights-based issues (particularly those concerning speech), and the potential for inconsistencies across the EU require further consideration.

The Brazilian model, with responsibilities attributed to CGI.br, is also formal and hierarchical. A long list of new attributions is provided in the Bill, with CGI.br's composition, general function, and structure set by a presidential decree. However, unlike ODS, CGI.br is not subject to a certification process. Due to its multi-stakeholder nature, it can also be considered to some extent a network-based model, where representatives in the committee represent wider sets of interested parties who provide input on platform regulation. While CGI.br has valuable subject-matter expertise, its limited resources and fragile legal status raise questions about its suitability for the proposed regulatory intermediary role in its current form. Addressing these weaknesses would be necessary for CGI.br to function effectively and fulfil the many functions currently proposed by Bill 2630/2020.

TABLE 3: Strengths and weaknesses of enhanced self-regulation in the EU DSA and in the Brazilian Bill

	EU DIGITAL SERVICES ACT	BRAZILIAN BILL 2630/2020
PRIMARY	Review of platform decisions by an	Supporting the development and

	EU DIGITAL SERVICES ACT	BRAZILIAN BILL 2630/2020
CONCERN	independent body, provides additional layer of appeal and redress	implementation of platforms' obligations
STRENGTHS IN ENHANCED SELF-REGULATORY MODELS	Clear requirements for impartiality and financial independence of ODS bodies, combined with market-based incentives fostering competition and potentially improving service quality.	CGI.br's long history in internet governance; greater expertise and many entry points across the content moderation cycle, with flexibility around policymaking, advising platforms, and enforcing rules.
WEAKNESSES IN ENHANCED SELF-REGULATORY MODELS	Lack of binding enforcement for ODS decisions. Focus on consensual solutions might not be well-suited for adjudicating complex rights-based disputes. Certification by national jurisdictions raises concerns about inconsistencies in quality and decision-making across the EU.	Concerns about CGI.br's ability to handle the workload and resource demands of a full-fledged regulatory body, particularly given its limited financial and human resources. The possibility of changes to CGI.br's composition and functions through decrees.

Ultimately, the effectiveness of content moderation systems on modern platforms hinges on their ability to balance power dynamics among states, technology companies, and the public. The analysis indicates that legal and regulatory design decisions (some made before the current wave of platform regulation, such as CGI.br's structuring) can significantly impact platform governance and ultimately influence users' experiences. Successfully incorporating regulatory intermediaries contributes to fulfilling the delicate balance that platform regulation requires, both through the diversity of voices they integrate and their ability to mediate relevant disputes.

As such, this article offers recommendations for the implementation phase of the DSA and informs ongoing legislative discussions in Brazil, aiding in the design of platform regulation rules. It also offers broader lessons that can contribute to the creation of more effective and democratic systems of online governance beyond these two cases, stressing the potential of regulatory intermediaries in this space.

ACKNOWLEDGEMENTS

I am grateful to Luke Richards for excellent research assistance, and to Octavio Vinhas and Beatriz Botelo Arcila for their constructive review and helpful suggestions. I also thank Anne Bellon and Frédéric Dubois for their careful editing and additional suggestions that further improved the manuscript. Any remaining errors are my own.

References

- Abbott, K. W., Levi-Faur, D., & Snidal, D. (2017a). Enriching the RIT framework. *The ANNALS of the American Academy of Political and Social Science*, 670(1), 280–288. <https://doi.org/10.1177/0002716217694593>
- Abbott, K. W., Levi-Faur, D., & Snidal, D. (2017b). Introducing regulatory intermediaries. *The ANNALS of the American Academy of Political and Social Science*, 670(1), 6–13. <https://doi.org/10.1177/0002716217695519>
- Araujo Monteiro Neto, J. (2018). *The operation of multistakeholderism in Brazilian internet governance: Governance innovation through multistakeholderism generativity* [University of Kent]. <https://kar.kent.ac.uk/76961/>
- Auld, G., & Renckens, S. (2017). Rule-making feedbacks through intermediation and evaluation in transnational private governance. *The ANNALS of the American Academy of Political and Social Science*, 670(1), 93–111. <https://doi.org/10.1177/0002716217690185>
- Baik, J. (Sophia), & Sridharan, H. (2024). Civil rights audits as counterpublic strategy: Articulating the responsibility and failure to care for marginalized communities in platform governance. *Information, Communication & Society*, 27(5), 836–855. <https://doi.org/10.1080/1369118X.2023.2227685>
- Bietti, E. (2023). A genealogy of digital platform regulation. *Georgetown Law Technology Review*, 7(1). <https://doi.org/10.2139/ssrn.3859487>
- Buiten, M. (2021). The Digital Services Act: From intermediary liability to platform regulation. *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, 123, 361–380. <https://doi.org/10.2139/ssrn.3876328>
- Caplan, R. (2023). Networked platform governance: The construction of the democratic platform. *International Journal of Communication*, 17, 22.
- Casarosa, F. (2023). Out-of-court dispute settlement mechanisms for failures in content moderation. *Journal of Intellectual Property, Information Technology, and Electronic Commerce Law*, 14(3), 391–402.
- Coimisiún Meán. (2024). Coimisiún na Meán certifies the first out-of-court dispute settlement body in Ireland. In *Coimisiún na Meán*. <https://www.cnam.ie/coimisiun-na-mean-certifies-the-first-out-of-court-dispute-settlement-body-in-ireland/>
- Douek, E. (2022). Content moderation as systems thinking. *Harvard Law Review*, 136(2), 526–607.
- Douek, E. (2024). The meta oversight board and the empty promise of legitimacy. *Harvard Journal of Law & Technology*, 37. <https://doi.org/10.2139/ssrn.4565180>
- Dvoskin, B. (2023). Expertise and participation in the facebook oversight board: From reason to will. *Telecommunications Policy*, 47(5). <https://doi.org/10.1016/j.telpol.2022.102463>
- Frosio, G. (2022). Regulatory shift in state intervention: From intermediary liability to responsibility. In E. Celeste, A. Heldt, & C. I. Keller (Eds.), *Constitutionalising Social Media* (pp. 151–176). Hart Publishing. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3850483
- Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press. <https://www.researchgate.net/publication/32718618>

2_Custodians_of_the_internet_Platforms_content_moderation_and_the_hidden_decisions_that_shape_social_media

Gillespie, T., Aufderheide, P., Carmi, E., Gerrard, Y., Gorwa, R., Matamoros-Fernández, A., Roberts, S. T., Sinnreich, A., & Myers West, S. (2020). Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates. *Internet Policy Review*, 9(4). <https://doi.org/10.14763/2020.4.1512>

Griffin, R. (2022). The sanitised platform. *JIPITEC – Journal of Intellectual Property, Information Technology, and E-Commerce Law*, 13, 36–52.

Griffin, R. (2023). Public and private power in social media governance: Multistakeholderism, the rule of law and democratic accountability. *Transnational Legal Theory*, 14(1), 46–89. <https://doi.org/10.1080/20414005.2023.2203538>

Hartmann, I. (2023, January 9). *In Brazil, the courts' efforts to curb disinformation had unintended consequences*. Centre for International Governance Information. <https://www.cigionline.org/articles/in-brazil-the-courts-efforts-to-curb-disinformation-had-unintended-consequences/>

Heldt, A., & Dreyer, S. (2021). Competent third parties and content moderation on platforms: Potentials of independent decision-making bodies from a governance structure perspective. *Journal of Information Policy*, 11, 266–300. <https://doi.org/10.5325/jinfopoli.11.2021.0266>

Howard, J. W., & Kira, B. (2024). *Remove or reduce: Demotion, content moderation, and human rights*. SSRN. <https://doi.org/10.2139/ssrn.4891835>

Husovec, Martin. (2023). *Rising above liability: The Digital Services Act as a blueprint for the second generation of global internet rules*. 38(3), 883–920. <https://doi.org/10.15779/Z38M902431>

Iglesias Keller, C., & Martins Dos Santos, B. (2024). Trackers and chasers. In C. Aguerre, M. Campbell-Verduyn, & J. A. Scholte, *Global digital data governance* (1st ed., pp. 148–164). Routledge. <https://doi.org/10.4324/9781003388418-11>

Jhaver, S., Frey, S., & Zhang, A. X. (2023). Decentralizing platform power: A design space of multi-level governance in online social platforms. *Social Media + Society*, 9(4). <https://doi.org/10.1177/20563051231207857>

Jost, I., & Cruz, F. B. (2024). X vs. Brazil: What questions remain after the turmoil? In *Tech Policy Press*. <https://www.techpolicy.press/x-vs-brazil-what-questions-remain-after-the-turmoil/>

Keller, C. I. (2022). The perks of co-regulation: An institutional arrangement for social media regulation? In E. Celeste, A. Heldt, & C. I. Keller (Eds.), *Constitutionalising social media* (pp. 217–234). Hart Publishing.

Kira, B. (2025). Inter-agency coordination and digital platform regulation: Lessons from the Whatsapp case in Brazil. *International Review of Law, Computers & Technology*, 39(1), 6–29. <https://doi.org/10.1080/13600869.2024.2351671>

Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131, 1598–1670.

Klonick, K. (2020). The Facebook oversight board: Creating an independent institution to adjudicate online free expression. *The Yale Law Journal*, 129(8). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3639234

Klonick, K. (2023). Of systems thinking and straw men. *Harvard Law Review Forum*, 136(6), 339–362.

Kruck, A. (2017). Asymmetry in empowering and disempowering private intermediaries: The case of credit rating agencies. *The ANNALS of the American Academy of Political and Social Science*, 670(1), 133–151. <https://doi.org/10.1177/0002716217691459>

Lytton, T. D. (2017). The taming of the stew: Regulatory intermediaries in food safety governance. *The ANNALS of the American Academy of Political and Social Science*, 670(1), 78–92. <https://doi.org/10.1177/0002716217690330>

Machado, C. C. V., & Aguiar, T. H. (2023). Emerging regulations on content moderation and misinformation policies of online media platforms: Accommodating the duty of care into intermediary liability models. *Business and Human Rights Journal*, 8(2), 244–251. <https://doi.org/10.1017/bhj.2023.25>

Marwick, A. E., & Caplan, R. (2018). Drinking male tears: Language, the manosphere, and networked harassment. *Feminist Media Studies*, 18(4), 543–559. <https://doi.org/10.1080/14680777.2018.1450568>

Medzini, R. (2021). Governing the shadow of hierarchy: Enhanced self-regulation in European data protection codes and certifications. *Internet Policy Review*, 10(3). <https://doi.org/10.14763/2021.3.1577>

Medzini, R. (2022). Enhanced self-regulation: The case of Facebook's content governance. *New Media & Society*, 24(10), 2227–2251. <https://doi.org/10.1177/1461444821989352>

Medzini, R., & Levi-Faur, D. (2023). Self-governance via intermediaries: Credibility in three different modes of governance. *Journal of Comparative Policy Analysis: Research and Practice*, 25(3), 323–345. <https://doi.org/10.1080/13876988.2022.2155516>

Ortolani, P. (2022). If you build it, they will come. *Verfassungsblog: On Matters Constitutional*. <http://doi.org/10.17176/20221107-095646-0>

Oversight Board. (2024, October 8). *Statements from the oversight board trust and oversight board members on the announcement of the Appeals Centre Europe*. <https://www.oversightboard.com/news/statements-from-the-oversight-board-trust-and-oversight-board-members-on-the-announcement-of-the-appeals-centre-europe/>

Roth, Y., & Lai, S. (2024). Securing federated platforms: Collective risks and responses. *Journal of Online Trust and Safety*, 2(2). <https://doi.org/10.54501/jots.v2i2.171>

Special Committee. (2023). *Parecer proferido em plenário ao projeto de Lei No 2.630, de 2020, e apêndices [Report of the Special Committee on Bill No. 2630/2020, and its appendices]*. Brazilian Chamber of Deputies. https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2265334

Uduba, S., Maronikolakis, A., & Wisiolek, A. (2023). Ethical scaling for content moderation: Extreme speech and the (in)significance of artificial intelligence. *Big Data & Society*, 10(1). <https://doi.org/10.1177/20539517231172424>

Vinhas, O., & Bastos, M. (2023). The WEIRD governance of fact-checking and the politics of content moderation. *New Media & Society*, 0(0). <https://doi.org/10.1177/14614448231213942>

Wimmers, J. (2021). The out-of-court dispute settlement mechanism in the Digital Services Act. *Journal of Intellectual Property, Information Technology, and Electronic Commerce Law*, 12(5), 381–401.

Woods, L. (2019). The duty of care in the Online Harms White Paper. *Journal of Media Law*, 11(1), 6–17. <https://doi.org/10.1080/17577632.2019.1668605>

Published by



in cooperation with

