INTERNET POLICY REVIEW

Volume 14 **Issue** 1

SPECIAL ISSUE

OPEN ACCESS

Introduction to the special issue on content moderation on digital platforms

Romain Badouard *Paris-Panthéon-Assas University* romain.badouard@u-paris2.fr
Anne Bellon *University of Technology of Compiègne*

DOI: https://doi.org/10.14763/2025.1.2005

Published: 31 March 2025
Received: 4 December 2024 **Accepted:** 19 December 2024

Funding: The authors did not receive any funding for this research.
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: Badouard, R., & Bellon, A. (2025). Introduction to the special issue on content moderation on digital platforms. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.2005

Keywords: Content moderation, Platform governance, Civil society, Digital platforms

Abstract: In this special issue, we refer to "content moderation" as the multi-dimensional process through which content produced by users is monitored, filtered, ordered, enhanced, monetised or deleted on social media platforms. This process encompasses a great diversity of actors who develop specific practices of content regulation. Users, non-governmental organisations (NGOs), activists, journalists, advertisers, experts, designers and researchers are becoming more and more involved in moderation-related activities, apart from, in partnership with, or against public authorities and firms. However, their precise contribution to the democratisation of content regulation, and to the balance between public and private interests in platform governance, remains little studied. Following the call to expand content moderation research beyond the relationship between states and firms (Gillespie et al., 2020), the goal of this special issue is to gather empirical studies that characterise the contribution of non-state actors in the current internet regulatory framework, and provide new insights on their various actions and strategies.

PAPERS IN THIS SPECIAL ISSUE

↓Introduction to the special issue on content moderation on digital platforms
Romain Badouard, *Université de Cergy-Pontoise*, *France*, *Romain.Badouard@u-cergy.fr*
Anne Bellon, *University of Technology of Compiègne*, *France*
Regulatory intermediaries in content moderation Beatriz Kira, *University of Sussex*
Article 22 Digital Services Act: Building trust with trusted flaggers Jacob van de Kerkhof, *Utrecht University*
The role of civil society organisations in co-regulating online hate speech in the EU: A bounded empowermentBarthélémy Michalon, *Sciences Po*
Framing the role of experts in platform governance: negotiating the code of practice on disinformation as a case study Kateryna Chystoforova, *European University Institute*
Urbano Reviglio, *European University Institute*
Aspirational platform governance: How creators legitimise content moderation through accusations of bias Blake Hallinan, *Hebrew University of Jerusalem*
CJ Reynolds, *Hebrew University of Jerusalem*
Yehonatan Kuperberg, *Hebrew University of Jerusalem*
Omer Rothenstein, *Hebrew University of Jerusalem*
Stop hate for profit: Evaluating the mobilisation of advertisers and the advertising industry to regulate content moderation on digital platforms Steph Hill, *University of Leicester*
Civil society's role in constitutionalising global content governance Nicola Palladino, *University of Salerno*
Dennis Redeker, *University of Bremen*
Edoardo Celeste, *Dublin City University*
Online advertising, content moderation, and corporate accountability: a civil society perspective Alex Rochefort, *Boston University*
The case for prosocial tech design governanceLisa Schirch, *University of Notre Dame*
The realm of digital content regulation as a social space: Sociogenesis of moderation norms and policies on Twitch platform Nathan Ferret, *ENS de Lyon*
Labour pains: Content moderation challenges in Mastodon growth Charlotte Spencer-Smith, *University of Salzburg*
Tales Tomaz, *University of Salzburg*
Safer spaces by design? Federated socio-technical architectures in content moderation Francesca Musiani, *National Centre for Scientific Research (CNRS)*
Ksenia Ermoshina, *National Centre for Scientific Research (CNRS)*

Introduction to the special issue on content moderation on digital platforms

In January 2025, Mark Zuckerberg announced his intention to relax moderation rules on his company's social media platforms and to end fact-checking partnerships with US media in favour of a community-driven system. His declaration raised major concern, as such moderation policy changes could allow hate speech and misinformation to gain new momentum on Facebook and Instagram. Implicit-

ly, this decision also demonstrates that moderation is a key dimension of services offered by digital platforms and that civil society organisations, at various levels, have consistently been involved in this process. What's more, Zuckerberg's statements and the debates they brought about show the extent to which content regulation, and the role played by tiered actors, has become a major political issue for democracies.

Indeed, facing a growing political demand to fight hate and disinformation online, European governments have launched regulatory initiatives, such as the NetzDG in Germany[1], the Law on fake news in France[2]or the European Digital Services Act[3]. This emerging regulatory framework has given way to a closer collaboration between states and firms. This evolution has transformed the power relation between public authorities and private platforms, and has shifted the attention of scholars on this dual relationship (Keller, 2022). Yet this antagonism only partially captures the governance of speech and information online, as it has recently evolved.

In this special issue, we refer to "content moderation" as the multi-dimensional process through which content produced by users is monitored, filtered, ordered, enhanced, monetised or deleted on social media platforms. This process encompasses a great diversity of actors who develop specific practices of content regulation. Users, non-governmental organisations (NGOs), activists, journalists, advertisers, experts, designers and researchers are becoming more and more involved in moderation-related activities, apart from, in partnership with, or against public authorities and firms. However, their precise contribution to the democratisation of content regulation, and to the balance between public and private interests in platform governance, remains little studied. Following the call to expand content moderation research beyond the relationship between states and firms (Gillespie et al., 2020), the goal of this special issue is to gather empirical studies that characterise the contribution of non-state actors in the current internet regulatory framework, and provide new insights on their various actions and strategies.

1. *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken* (Law to improve law enforcement in social networks) *or Netzwerkdurchsetzungsgesetz – NetzDG*, September 2017: https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html
2. *LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information* (Law against the manipulation of information): https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559/
3. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act): https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065

Two distinct approaches are pursued in this special issue. The first one is "institutional" and rooted in legal and political scholarship: it focuses on the institutionalisation of civil society participation in content regulation processes, analysing the structures, scope, and limitations that underpin civil society contribution. A second approach that primarily considers "alternative" governance mechanisms is more prominently embedded in social science research. It acknowledges the limits of multistakeholderism and concentrates on third-party actors that challenge the moderation of commercial platforms. The objective of this special issue is thus twofold: to specify the various types of relationships at play between civil society organisations (CSOs), platforms, and public authorities, elucidating their strengths and limitations, while simultaneously to analyse creative governance models and principles developed within civil society, with the aim of conceptualising the conditions for their expansion.

This special issue gathers twelve contributions, consisting of ten research articles and two opinion pieces which focus on diverse case studies and are rooted in varied disciplinary traditions (law, political science, sociology, science and technology studies, communication and media studies). The issue examines a wide array of civil society actors, with distinct means of action, and involved in diverse symbolic and geographical spaces. However, a common inquiry spans through these contributions: under what conditions can content regulation truly be fair, effective and inclusive? In the introduction to this special issue, we suggest several avenues for reflection to address this question, drawing from a (non-exhaustive) state of the art on the subject. We first present the various models of governance underpinning content regulation, and then we discuss the category "civil society" and what it covers in terms of social groups and possibilities for action. Finally, we emphasise the importance of studying civil society initiatives to devise new forms of content governance in the face of increasingly worrying partnerships between businesses and governments.

From the state-firm relationship to the global governance of content regulation

In the 1990s, the expansion of the internet as a critical infrastructure for many aspects of social and economic exchanges raised the fundamental question of who should govern this global information network. The resulting internet governance emerged as a compromise between two different cultures of governance: bottom-up informal processes of internet communities and the private sector on the one hand, top-down formal approaches of governments and intergovernmental organi-

sations on the other. The United Nations notably recognised the role of tech communities in setting principles and standards for the internet through bottom-up cooperation and "multistakeholder" involvement (Cammaerts, 2011; Raymond & DeNardis, 2016). Multistakeholderism came to define a preferred governance arrangement bound to ensure a dialogue between governments, firms and civil society in governing the internet. From the World Summit on the Information Society to the Internet Governance Forum, a rich literature has then focused on and discussed the contribution of civil society organisations in shaping internet governance (Bygrave & Bing; 2009; DeNardis, 2014; Radu, 2019). Yet, this bulk of literature has mostly paid attention to the governance of the global infrastructure, the management of technical resources (such as IP addresses and domain names) or the protection of digital rights and, much less to content-related issues (Badouard, 2020).

Over the past ten years, scholars of content regulation have started to assess the role of third parties in the global governance of speech and information online, reviving interest in the idea of multistakeholder governance. If recent content regulation initiatives bear the legacy of internet governance multistakeholder experimentations, they also convey the specific nature of this political issue, which is seen by some as dominated by national governments – information control having been a core state prerogative over centuries – and large commercial platforms – as the main infrastructure behind content circulation online. The growing interest in "middle-level governance" (Jhaver et al., 2023) or alternative arrangements between formal government control and private moderation addresses the crisis of trust surrounding commercial moderation (Zuckerman & Rajendra-Nicolucci, 2023) and the risk of state censorship (Balkin, 2017). Drawing on the conceptual model of the "governance triangle" by Abbott and Snidal (2009), Robert Gorwa (2019) has highlighted the role of "non-governmental organizations", both as independent players and partners to state and firms in content regulation. He emphasises a shift from traditional command-and-control regulation towards multiple and overlapping forms of multistakeholder arrangements that shape the governance of social media platforms and the debates on content moderation.

Nevertheless, partnerships between civil society, platforms, and public authorities do not necessarily translate into an empowerment of citizens and CSOs that would mechanically entail a greater balance between public and private interests in the governance of online speech. Formal partnerships frame and constrain the means of action available to civil society, without always allowing associations or users to transcend, negotiate or challenge these frameworks. Regardless of the forms of

power delegation they benefit from (such as defining moderation standards, implementing publication policies, or evaluating procedures), civil society finds its power curtailed by states and platforms that determine who participates, for what purpose, and to what end.

The academic literature on the subject, as reflected by the contributions gathered in this issue, oscillates between acknowledging procedural creativity in the manner of opening governance mechanisms to new actors, and recognising the relative failure of a genuine democratisation of online content regulation. As a much quoted reference in the papers, Rotem Medzini (2021, 2022) discusses in his works how content regulation has evolved from self-regulation to polycentric governance regimes, with an increased role for "regulatory intermediaries" – that is, non-state and non-corporate actors that constrain, shape, and guide moderation practices of large platforms. However, he simultaneously observes that these new regimes only produce a form of "enhanced self-regulation" for platforms (Medzini, 2022), improving their self-regulatory capacities without engendering a real sharing of power.

In a similar vein, Robyn Caplan (2023) has employed the concept of "networked governance" to describe how a broad array of actors external to platforms contribute to the creation, implementation, and legitimation of private moderation standards. Nevertheless, she also observes that these partnerships can lead to forms of instrumentalisation of civil society actions, generating legitimacy for the platforms while fostering distrust among the organisations and citizens participating in these novel forms of regulation.

Separating themselves from the concept of multistakeholderism as an idealised (Epstein, 2011) or mythical (Dany & Freistein, 2016) acceptance of civil society participation, these works remind us that civil society's involvement does not operate in a vacuum, but is embedded in a set of social, political and economic relationships organising the production of regulation. Studying the complex interactions between state, platforms and other actors involved in content regulation, the aforementioned scholars provide a conceptual grid to analyse how power distribution between various stakeholders can inform the varying outcomes of different governance regimes. Notably, power asymmetries between users and platforms (Nielsen & Ganter, 2022), regarding access to information, financial and technical resources, inform the ability of non-state actors to influence or contest the transformation of content regulation. This issue articulates this analytical framework with various case studies to expand our understanding of civil society participation in content regulation and suggest possible improvements for a more inclusive

and democratic governance of our information and communication infrastructures.

Defining civil society organisations

As a global and unique entity, civil society is an "elusive" concept (Mouffe, 2010), a loose agglomeration of diverse actors with diverse views on internet related issues (Mueller et al., 2007). From informal collectives of users to highly institutionalised entities, civil society is a broad and often ill-defined category, sometimes instrumentalised by firms and governments alike to justify key decisions or advance their own interests in a multistakeholder governance framework. The category includes a wide range of groups, with various and sometimes conflicting interests and resources such as advocacy coalitions, international non-governmental organisations, academic researchers, activist investors, experts, journalists and even individual users. Thus, the first challenge of this special issue is to take a closer look at this very diverse group of organisations, to establish heuristic distinctions between them, and to understand what dynamics inform their relationships. Indeed, the focus on Transnational Advocacy Networks in multistakeholder analysis and the assertion that they share principles and values (Keck & Sikkink, 2014) has often led empirical studies to neglect the conflicts and differences between the goals and perceptions of organisations within and outside these networks. A key limitation of the concept of multistakeholderism in general has been to overlook the huge disparities in resources and influence but also special interest and agenda among civil society stakeholders. Reinstating this diversity and the power structure underlying various stakeholders' participation is thus both an academic and a democratic issue, as the pretense of embodying civil society through a unique set of universal principles is used by corporate powers to legitimate their moderation activities, as suggested and exemplified by the Facebook Oversight Board (Dvoskin, 2023).

Common distinctions established in the literature often focus on the structure of the organisations – either informal or highly institutionalised –, their scope of intervention – either local or transnational – and their main role – from expertise to coordination (e.g. Tjahja, Meyer, & Shahin, 2021). In addition, new regulatory arrangements are based on some forms of accreditation or contract, granted either by government or by the platforms themselves that associate third-party partners in the monitoring or auditing of content moderation. The attribution and content of contracts or state-supported partnerships thus become key in understanding the participation of various organisations. As a related issue, the unequal distribution of political, social and financial resources among civil society organisations can lead to the potential exclusion or under-representation of minority groups or less-

endowed organisations in content regulation. Studying formal participation at international forums shows for example the over-representation of organisations from the Global North (Grover, 2022). The dominance of some civil society actors in the global governance of the internet has eventually led to the implementation of mechanisms for civil society's legitimate representation and involvement in decision-making. Yet greater inclusion often comes at the cost of cohesion, inducing coordination costs for large transnational networks (Haristya, 2020). So, there's not always an equivalence between a player's size, its institutional legitimacy and its real impact on the definition of regulation.

Unpacking the meaning of civil society participation also implies a deeper analysis of the various contributions and roles played by non-governmental organisations in shaping the global governance of speech and information online. As a second main challenge of this special issue, we consider that the question of civil society participation should not be reduced to their effective contribution to the definition of legal standards. The aim is to move away from the evaluation of influence alone, and take a broader look at the diversity of political roles assumed by civil society organisations and individuals in shaping online moderation. In the following section, we distinguish between four main roles associated with civil society participation in the governance of online content: *direct or undirect moderation, expertise, critique and norm settings*.

Various functions, limited power

Direct and indirect contribution to moderation

First of all, non-state and non-platforms actors have long exerted a direct role in moderating content on the internet and the web. In smaller online communities such as Usenet groups, personal blogs or mailing lists, content moderation was in the hands of the community that dealt on a daily basis with the enforcement of community standards (Postigo, 2009; Paloque-Bergès, 2018), through the removal of offensive posts or the banning of users with aggressive behaviour, for example. Moderation then took many forms, from the sole intervention of a webmaster or a community manager in the discussion to public arbitration and *ad hoc* councils. Such moderation persists, notably on decentralised networks, such as Mastodon or Matrix/Element (as studied in this issue), using the federated architecture to promote community-based moderation on large social media. While most social media platforms have grown to abandon this form of purely community-based moderation, it continues at a smaller level with volunteer moderation being the main rule on Facebook Groups or subreddits for example, where moderators can order

posts, filter words and ban users. Besides, platform users are still associated in many ways to the detection of illicit content and their deletion, inviting us to conceptualise “platforms and users as partners in regulation rather than as subjects” (Helberger et al., 2018). First, through their stated preferences, individual users can mute content they dislike, operating a form of personalised moderation that contributes to changing the flow of information. Second, platforms receive billions of flags and reports from users about illegal, abusive or unwanted content. Third, in 2021, Twitter/X launched a crowdsourced fact-checking programme that allows users to add contextual notes to posts and votes for the most helpful notes that will be publicly shown. Although criticised for its limited effects and potential misuse, this form of community-based fact-checking has recently been endorsed by Mark Zuckerberg to replace formal partnerships with US media outlets on Meta’s social networks.

By enabling users to report content they deem in violation of publication standards or national laws, content flagging addresses two fundamental needs of platforms (Crawford & Gillespie, 2016): firstly, to enhance moderation efficiency by leveraging large-scale human resources in close proximity to the content; and secondly, to legitimise moderation by justifying certain decisions through a process that allegedly represent the views of the “community”. However, the true impact of flagging on content moderation is far more ambiguous. Beyond its strategic use by political activists for censoring opponents (Badouard, 2020) or digital vigilantism (Loveluck, 2016), a 2015 study highlighted that on pre-Musk Twitter/X, nearly 50% of reports constituted abusive flagging (Matias et al., 2015). Similarly, YouTube's reports on NetzDG enforcement in Germany showed that out of over 190,000 reports made in the first half of 2023 concerning law enforcement, only slightly more than 30,000 resulted in content removal (approximately 16%). In the face of widespread automated moderation tools (Gorwa, Binns, & Katzenbach, 2020; Badouard, 2021) flagging ultimately represents only a small portion of moderation actions undertaken by platforms. Ultimately, flagging proves to be more of a customer relations service (Gillespie, 2018), allowing users to express dissatisfaction through a dedicated mechanism, than an effective moderation tool.

To address concern on the efficiency and legitimacy of flagging, platforms and states have promoted "superflagger" systems, offering specific reporting channels to third-party organisations to prioritise their claims over other members (Gillespie, 2017). This system has been reinforced by the implementation of the Digital Services Act, whose Article 22 has provisioned to extend and institutionalise this cooperation mechanism with the formal designation of "trusted flaggers" whose

notifications must be treated with priority by the platform. In practice, however, priority flagging mechanisms demonstrate several limitations. Beyond the voluntary and potentially traumatic work provided by these partnerships (Gillespie, 2018), the partner selection process remains opaque and entirely controlled by platforms or public authorities. CSOs are confined to a specific role from which they cannot extricate themselves, as highlighted in Barthélémy Michalon's article in this issue (Michalon, 2025). Worse still, these mechanisms can be subverted by states or the platforms themselves. In this issue, Jacob van de Kerkhof (van de Kerkhof, 2025) demonstrates how institutions could access trusted flagger mechanisms to censor "lawful but awful" content, thereby circumventing national laws by removing content that is not illegal but contravenes platform publication standards, which are generally more restrictive than laws. The Digital Services Act brings transparency to the recruitment process by imposing selection criteria for trusted flaggers, such as expertise, independence, diligence, objectivity, and accuracy. However, many questions remain about how to evaluate these criteria in practice and the effective contribution of trusted flaggers: who will flag the trusted flaggers' abuses? Besides, as selection procedures are not harmonised at the European level, platforms retain great control over this process.

In a paper that was just published in this journal, Elizabeth Farries and Eugenia Siapera place the powerlessness of civil society within a broader neoliberal paradigm (Siapera & Farries, 2025). Delegating functions to third-party organisations does not necessarily mean delegating power to them, and the forms of multistakeholderism they study in the broader field of digital rights result in a strengthening of the power of platforms and public authorities. In this issue, Beatriz Kira also acknowledges the relative failure of participatory co-governance approaches so far, and calls for the establishment of hybrid regulatory systems (Kira, 2025). Within these systems, civil society organisations would engage *alongside* rather than *for* platforms, under the supervision of public authorities. The role of these authorities would be to ensure a real empowerment of CSOs and the transparency of partnerships with platforms.

Expertise

Given the complexity of socio-technical arrangements informing the global functioning of the internet and the web, expertise is a key asset in the implementation and evaluation of content regulation mechanisms, and internet governance in general (An & Yoo, 2019; Mărcuț, 2020). Experts have been associated at various stages of the policy-making process, especially at the EU level, from issue framing to enforcement. For example, the European Commission set up a high-level group

of experts as early as June 2018 to advise on policy initiatives to counter fake news and disinformation online. The Commission also promotes shared responsibility for the audit of platforms and the enforcement of new policies through the provision of expansive access to data for researchers. These mechanisms mirror more broadly the increased resort to expertise in risk regulation and policymaking, in domains as different as finance, pharmaceuticals, nuclear energy, immigration, etc. (e.g. Boswell, 2008; Demortain, 2023; Christensen, 2021). These works show that expertise is increasingly embedded in audit and risk-assessment practices. With regard to content regulation, expertise refers to various forms of knowledge, from legal analysis to design and computer science. It is thus a hybrid science used in regulation to establish the risks and benefits associated with content circulation and moderation, a form of knowledge produced by various social groups, stemming either from the academic field, corporate or non-profit organisations. Although most scholars engaged in policy-making and regulation embrace ideals of objectivity, transparency and evidence-based decision-making, the study of expertise production in content governance is deeply connected with the risk of "regulatory capture" (Stigler, 1971; Carpenter & Moss, 2014). With information asymmetries to the benefit of platforms and revolving door between expert groups and corporate laboratories, the autonomy of expertise has often been questioned and knowledge production has become a new battleground between private, public and independent experts for the governance of speech and information online.

In this issue, the mechanism of regulatory capture through science is notably highlighted by the contribution of Kateryna Chystoforova and Urban Reviglio (Chystoforova & Reviglio, 2025): they study the framework of partnerships with experts mobilised to combat disinformation at the European level, and emphasise how the work of these experts is constrained by a number of external factors, such as the selection process (which is controlled by platforms and public authorities), the data provided (which shapes the production of results), and finally the consideration of expert findings (as experts have no power over how their findings are translated into policies). The expert groups themselves can also be infiltrated by platform companies that send data scientists or lawyers to act as lobbyists in expert groups. The mechanisms of "soft corruption," well-documented in other areas of lobbying, remain underexplored in the context of expertise related to platform governance, as Chystoforova and Reviglio point out, calling for greater transparency in the expert selection and contribution processes.

In light of this situation, experts may seek to transcend the confines of partnerships to assert their independence. Barthélémy Michalon, in this issue, illustrates

how CSOs engaged in partnerships with the European Commission to audit codes of conduct established by private actors, are gradually emancipating themselves from the roles assigned to them in order to produce an independent evaluation (Michalon, 2025). By creating their own spaces of exchanges where they share their expertise and by relying on support from their members and the general public, they are better positioned to influence the Commission's decisions, even though their role remains limited to evaluating *a posteriori* the code of conduct rather than defining new standards in this area.

Critique and norm-setting

Finally, civil society organisations play a vital role to enhance corporate accountability. They have a watchdog function through their advocacy, research, and investigations into platform practices (Gorwa, 2019). From individual whistleblowers – such as Frances Haugen disclosing the Facebook Files – to collective reports by academics, civil society groups have put forth the shortcomings of current content regulation regimes, their biases and potential chilling effect. Several examples are to be found in the literature, from breastfeeding movements in the US and more broadly the representation of nudity (Myers-West, 2017; Gillespie, 2018; Sybert, 2022), child safety associations (Gorwa, 2014), to LGBTQ communities and ethnic minorities (Nakamura, 2015; Grison et al., 2023), of activists who exposed and criticised potential dangers, discriminations and visibility reduction associated with social media moderation (or lack of). Creators and tiered industry actors – such as advertisers (Griffin, 2023a) – have also campaigned, more or less effectively, to push platform operators to be more active in their fight against hate online, by modifying their moderation policies as well as their content monetisation strategies.

The power of civil society in this domain stems from its ability to mobilise citizens and attract media attention to exert pressure on platforms from the outside. The contribution of CSOs in this area is recognised and sometimes pushes platforms or public authorities to develop their own strategies to circumvent or instrumentalise this power. In the realm of "brand safety," as previously mentioned, Steph Hill examines in this issue the Stop Hate for Profit campaign, which rallied advertisers to pressure Facebook into demonetising pages and groups identified as promoting hate speech (Hill, 2025). While the campaign's organisers declare it a success, notably achieving a civil rights audit of the platform and encouraging the "deplatforming" of alt-right movements, Facebook also established a partnership with a competing organisation, the Global Alliance for Responsible Media. According to Hill, this alternate partnership served as a shield for Facebook to avoid complying

with the more radical demands of the Stop Hate for Profit campaign. In this context, the partnership also acts as a means to defuse the mobilising strength of civil society by limiting the potential impacts that broader campaigns could have on public opinion.

The interplay between mobilisation actions and their impact on public opinion is thus central to assess the power of civil society. Social campaigns have indeed contributed to a "moral panic" (Cohen, 1972), as well as the emergence and recognition of new public issues regarding the protection of speech and the quality of information online. As an early case of content regulation debate, during the "copyright wars" (Patry, 2009) of the early 2000s, civil society organisations, notably from the tech community, have developed a new repertoire of action to fight the reinforcement of intellectual property online (Breindl & Briatte, 2013). More recently, the activism of civil society - such as child safety associations or civil rights organisations - has been key in building a new political demand for regulatory change in most western countries, and in sustaining the legitimacy of government intervention in this domain. After a decade of celebrating the democratic virtues, emancipatory effects and socio-economic gains associated with the internet, social movements contribute to a new critique of digital capitalism (Alexandre et al., 2022), assessing its impact on democratic values and human rights. Regarding new institutional arrangements, digital activists have also expressed concerns about the check and balances between private and public power. More generally, they often discuss and criticise the role of governments in promoting new forms of private censorship, in the name of the fight against terrorism or in order to protect intellectual property rights (see Quadrature du net, 2020; EFF, 2018).

As an address to these governance issues surrounding the free flow of information and the safeguard of the freedom of expression, civil society organisations put forth sets of principles to spur public discussion and affect the governance landscape. For example, in 2014 and 2015, a group of digital rights organisations developed the 'Manila Principles on Intermediary Liability', which recognises that intermediaries (internet service providers (ISPs), social networks, and hosting platforms) should not have legal liability over third-party content, and that content restriction requests should be transparent and respect due process. In 2018, a small group of civil society organisations and researchers, including the EFF, proposed the 'Santa Clara Principles for Content Moderation' (SCPs), advocating for specific recommendations on how companies should meet basic best practices for appeals, user notice, and transparency in their moderation processes. Through advocacy coalitions and their participation in large forums about the fight against hate

speech and disinformation, civil society organisations thus try to influence the norms and play their part in the definition of new regulatory arrangements.

In this special issue, several contributions illustrate how users or civil society organisations engage in these normative activities. In his opinion piece, Alex Rochefort recounts the work carried out by the organisation Ranking Digital Rights, which produces its own indicators of platforms' respect for human rights (Rochefort, 2025). The organisation's work thus offers open standards and data for evaluating platform behaviours in this area, following a logic of "statactivism" (Bruno, Didier, & Prévieux, 2014), in order to objectify a public problem and produce "the authority of facts." Lisa Schirch, for her part, reports on the activism at work in the field of ethical design, where designers concerned with developing technologies that respect users' attention and rights share applications and software to demonstrate that an alternative design, described as "pro-social," is possible on the internet (Schirch, 2025). Through app stores and other sharing spaces between designers, these activists contribute to the establishment and circulation of new design standards. At the intersection of these two approaches (digital rights and platform design), Nicola Palladino, Denis Redeker, and Edoardo Celeste examine the work of transnational advocacy networks (TANs) in the field of the digital "constitutionalisation" of content regulation (Palladino et al., 2025). By studying charters and declarations of rights in the domain of moderation, they demonstrate how these TANs contribute to the emergence of a set of normative values on how social media platforms ought to be governed, and how, through their practices of information sharing and exchange of best practices, they can facilitate the development of global standards for digital governance.

The creative power of civil society

Last but not least, one of the major strengths of civil society in online content governance is its ability to provide alternative solutions to commercial platforms and their moderation models. Regarding this aspect, we are witnessing a revival of interest among researchers for socio-historical approaches that study moderation practices that pre-existed the emergence of large commercial platforms in the 2000s, to inform current debates on content regulation (Zuckerman & Rajendra-Nicolucci, 2023). In these studies based on historical material one of the primary sources of inspiration pertains to the decentralised model of moderation in online forums. More generally, a significant portion of academic debate deals with decentralised architectures as a way to provide alternative moderation models (Bodó et al., 2021; Musiani & Ermoshina, 2022), particularly through case studies examin-

ing platforms such as Mastodon or BlueSky. The platforms of the Fediverse, of which Mastodon is the main representative, propose for instance a model wherein independent communities, each with their own moderation rules, federate with one another to create a larger public sphere of information sharing and debating. Within this space, internet users circulate and decide which types of content they wish to be exposed to. This model has the merit of combining community self-management with user decision-making power (through customisation), as demonstrated by Ksenia Ermoshina and Francesca Musiani, as well as Charlotte Spencer-Smith and Tales Tomaz, in this issue (Ermoshina & Musiani, 2025; Spencer-Smith & Tomaz, 2025). Beyond the power delegated to users, the decentralised model also has the collateral effect of making users aware of their belonging to the community and of the challenges of content governance (Mansoux & Roscam, 2020).

Yet, the decentralised model also exhibits a number of limitations. A significant challenge for decentralised social networks is the question of scalability. In this issue, Charlotte Spencer-Smith and Tales Tomaz demonstrate the difficulties encountered by moderators of Mastodon instances in coping with the arrival of new users following Elon Musk's acquisition of Twitter (Spencer-Smith & Tomaz, 2025). These difficulties are attributed not only to a lack of resources, time, or qualification but also to the cultural codes of newcomers (whether geographical - between global north and global south - or sociological codes), which often diverge from those of decentralised social network pioneers. The result is a conflict of norms, which engenders inclusivity challenges on these networks and renders the application of such a model improbable at the scale of large commercial platforms.

To move beyond naive or idealised visions of decentralisation, as the authors of this issue invite us to do, a more productive avenue may be to consider how centralised and decentralised models can interact within the same platform, application or network of websites. This articulation can be conceptualised in both "horizontal" and "vertical" ways (Jhaver et al., 2023): horizontal, in the sense of how decentralised networked communities can also benefit from centralised resources, particularly in the domain of moderation (automated detection tools, user registries, etc.). Vertical, in terms of how communities interact with higher levels of authority, as it is already the case within numerous platforms such as Reddit, Twitch or YouTube. Smaller communities on these platforms enjoy the power of self-regulation, establishing their own publication rules and their enforcement, while adhering to a common set of norms shared by all communities, the implementation of which is guaranteed by the platform. This interweaving of governance scales and tools, for which various degrees and conditions of autonomy

must be considered, outlines promising alternative moderation approaches, as evidenced by several articles in this issue.

The case of Twitch, studied by Nathan Ferret, is representative of this interweaving of governance scales, as the platform combines the centralisation of common rules and moderation tools with the distribution of moderation responsibility to participative self-managed channels (Ferret, 2025). Beyond the articulation of scales, Ferret focuses on those who perform moderation. Factors of engagement in moderation activities have generally been apprehended through a psychological lens (Bateman et al., 2021) linking involvement in moderation activity to attachment to a channel, a sense of obligation or reciprocity towards the community, or social recognition and symbolic retribution. Ferret rather demonstrates that these analyses underestimate the socialisation processes within these communities, partly determined by socio-demographic factors, such as the fact of finding oneself with peers who resemble us and with whom we share daily experiences.

Intermediate levels of governance can also protest against top-down modes of regulation entailing new forms of online discrimination. Internet users, and notably influencers, who often act as moderators at intermediate levels, can turn into norm entrepreneurs, as shown by Blake Hallinan, CJ Reynolds, Yehonatan Kuperberg, and Omer Rothenstein in this issue, through the case of "react videos." (Hallinan et al., 2025) On YouTube, content producers with high visibility question the platform's moderation policies and raise awareness within their community on potential discrimination or moderation abuse, effectively generating a form of "user-generated accountability" (Reynolds & Hallinan, 2024). Hallinan and colleagues describe these practices as "informal and cultural modes of platform governance". Although their direct influence on platform policies is difficult to assess, such practices tend to increase collective knowledge of moderation mechanisms and strengthen public scrutiny of moderators' work. Such community dynamics have been observed in the past among certain ethnic, religious, or gender minorities groups who mobilised against what they perceive as a form of discriminatory censorship by platforms (Nakamura, 2015; Grison et al., 2023). These mobilisations incorporate a productive dimension, where community members engage in evaluation and audit activities of moderation policies and develop technologies to fuel common knowledge and allow vulnerable users to extricate themselves from the power of platforms. Generally speaking, practices of "algorithmic resistance," even if not related to moderation, are based on the collective production of independent expertise (Bonini & Treré, 2024). However, this counter-power and the dissemination of independent practices remains directly linked to the visibility and

technical resources that are owned by or lacking to groups of users.

Producing alternative spaces for information and debate is therefore not only about building innovative technologies to be implemented in large commercial platforms. It is also about (re)shaping the social network ecosystem, the power relations that structure it, and the actors who animate it, towards the common good and general interest. This reorientation of the platform ecosystem can be achieved through more comprehensive regulation of the economic markets in which platforms operate (Griffin, 2023b), as well as through the creation of public service platforms. In Asia, the United States, and Europe, organisations and citizens are associating around the promotion of "digital public infrastructures" (Zuckerman, 2020). These infrastructures lean on a combination of technical protocols, design and political institutions, that inform the development of not-for-profit exchange spaces with a more transparent and participatory governance. For example, the Netherlands is developing the PubHub project, aiming to promote civic debate platforms on a national scale; the United States is experimenting with networked micro-spaces of conversation with Small Town; Taiwan is promoting participatory online governance engineering in the wake of Audrey Tang's activism; in France, the Conseil National du Numérique (National Digital Council) is campaigning for the promotion of communication protocols allowing unbundling and interoperability of social network services, to prepare for the entry of online sociability into the era of artificial intelligence. This new generation of social networks, emerging on the margins of commercial web spaces, is driven by the dynamism and creativity of civil society organisations, which, in this domain as in others, remain the main engine of social change.

Conclusion

This issue is structured around three main sections. The first encompasses four research articles examining formal partnerships between public institutions and civil society, identifying their strengths and weaknesses through the cases of regulatory intermediaries (Beatriz Kira), Trusted Flaggers (Jacob van de Kerkhof; Bathélémy Michalon), and experts in the field of disinformation (Kateryna Chystoforova & Urbano Reviglio). The second section focuses on various forms of activism adopted by users and civil society organisations (CSOs) to influence platform policies or produce alternatives. Notable case studies include react videos on YouTube (Blake Hallinan, CJ Reynolds, Yehonatan Kuperberg & Omer Rothenstein), campaigns in the domain of brand safety (Steph Hill), and the dynamics of constitutionalising content regulation (Nicola Palladino, Dennis Redeker & Edoardo Celeste). In addi-

tion to these four articles, two opinion pieces are included, addressing the production of independent assessments in the field of human rights compliance by platforms (David Rochefort), and activism in the domain of ethical design (Lisa Schirch). The final section concentrates on users' own contributions to online content regulation, examining the cases of Twitch (Nathan Ferret), Mastodon (Charlotte Spencer-Smith & Tales Tomaz), and the governance of federated infrastructures (Ksenia Ermoshina & Francesca Musiani).

The contributions to this special issue, in their diversity, invite us to rethink content regulation not merely as a technical or legal challenge but as a political and social project. While partnerships between platforms, civil society, and public authorities have recently been strengthened, broadening the scope of contributions to the study of platform governance,these various frameworks also reveal persistent power asymmetries among stakeholders, which constitute structural barriers to the true democratisation of content regulation. Civil society participation remains constrained by the parameters set by platforms and public authorities, limiting its impact on prevailing norms and practices.

Nevertheless, the emergence of alternative models – whether federated technical infrastructures, collaborative moderation practices, or innovative governance structures – offers promising avenues for rethinking content regulation beyond the models promoted by commercial platforms. These alternatives underscore the creative and transformative role that civil society can play in redirecting digital spaces toward the public interest and the common good.

Among the fruitful avenues for reflection suggested by this special issue is the future exploration of hybrid models capable of combining centralisation and decentralisation on a large scale, thereby reconciling efficiency and inclusion. Another promising direction involves examining how civil society actors can be integrated into moderation processes in countries of the Global South (Bouquillion et al., 2024), as most field studies still focus on Northern contexts. Exploring cultural publication and moderation practices in these regions could shed new light on the global dynamics of the digital public sphere.

This special issue thus serves as an invitation to explore new research directions, conduct more field studies, and move beyond existing conceptual frameworks to rethink the future of our information infrastructures and digital debate architectures.

References

Abbott, K. W., & Snidal, D. (2009). International regulation without international government: Improving IO performance through orchestration. *The Review of International Organizations*, *5*(3), 315–344. https://doi.org/10.1007/s11558-010-9092-3

Alexandre, O., Beuscart, J. S., & Broca, S. (2022). Une sociohistoire des critiques numériques [A sociohistory of digital critique]. *Réseaux*, *231*(1), 9–37. https://doi.org/10.3917/res.231.0009

An, J., & Yoo, I. T. (2019). Internet governance regimes by epistemic community: Formation and diffusion in Asia. *Global Governance*, *25*(1), 123–148. https://doi.org/10.1163/19426720-02501008

Badouard, R. (2020). *Les nouvelles lois du web: Modération et censure [The new laws of the web: Moderation and censorship]*. Seuil.

Badouard, R. (2021). Moderation on social networks: The policies of platforms faced with new content regulation obligations in Europe. *Réseaux*, *225*(1), 87–120. https://doi.org/10.3917/res.225.0087

Balkin, J. M. (2017). Free speech in the algorithmic society: Big data, private governance, and new school speech regulation. *UC Davis Law Review*, *51*(4), 1149–1208.

Bateman, P. J., Gray, P. H., & Butler, B. S. (2011). Research note—The impact of community commitment on participation in online communities. *Information Systems Research*, *22*(4), 841–854. https://doi.org/10.1287/isre.1090.0265

Bodó, B., Brekke, J. K., & Hoepman, J.-H. (2021). Decentralisation: A multidisciplinary perspective. *Internet Policy Review*, *10*(2). https://doi.org/10.14763/2021.2.1563

Bonini, T., & Treré, E. (2024). *Algorithms of resistance: The everyday fight against platform power*. The MIT Press. https://direct.mit.edu/books/book/5721/Algorithms-of-ResistanceThe-Everyday-Fight-against

Boswell, C. (2008). The political functions of expert knowledge: Knowledge and legitimation in European Union immigration policy. *Journal of European Public Policy*, *15*(4), 471–488. https://doi.org/10.1080/13501760801996634

Bouquillion, P., Ithurbide, C., & Mattelart, T. (2023). *Digital platforms and the global south: Reconfiguring power relations in the cultural industries* (1st ed.). Routledge. https://doi.org/10.4324/9781003391746

Breindl, Y., & Briatte, F. (2013). Digital protest skills and online activism against copyright reform in France and the European Union. *Policy & Internet*, *5*(1), 27–55. https://doi.org/10.1002/poi3.21

Bruno, I., Didier, E., & Prévieux, J. (2014). *Statactivisme. Comment lutter avec des nombres [Statactivism: How to fight with numbers]*. Éditions La Découverte.

Bygrave, L. A., & Bing, J. (Eds.). (2009). *Internet governance: Infrastructure and institutions*. Oxford University Press.

Cammaerts, B. (2011). Power dynamics in multi-stakeholder policy processes and intra-civil society networking. In R. Mansell & M. Raboy (Eds.), *The handbook of global media and communication policy* (pp. 129–146). Wiley-Blackwell.

Caplan, R. (2023). Networked platform governance: The construction of the democratic platform. *International Journal of Communication*, *17*, 3451–3472.

Carpenter, D., & Moss, D. A. (Eds.). (2013). *Preventing regulatory capture: Special interest influence and how to limit it*. Cambridge University Press.

Christensen, J. (2021). Expert knowledge and policymaking: A multi-disciplinary research agenda. *Policy & Politics*, *49*(3), 455–471. https://doi.org/10.1332/030557320X15898190680037

Chystoforova, K., & Reviglio, U. (2025). Framing the role of experts in platform governance: Negotiating the code of practice on disinformation as a case study. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.1823

Cohen, S. (1972). *Folk devils and moral panics: The creation of the Mods and Rockers*. Martin Robertson.

Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society*, *18*(3), 410–428. https://doi.org/10.1177/1461444814543163

Dany, C., & Freistein, K. (2016). Global governance and the myth of civil society participation. In *Myth and narrative in international politics: Interpretive approaches to the study of IR* (pp. 229–248).

Demortain, D. (2023). Experts in the regulation of technology and risk: An ecological perspective on regulatory science. In *The Oxford handbook of expertise and democratic politics* (pp. 282–313). Oxford University Press.

DeNardis, L. (2014). *The global war for internet governance*. Yale University Press.

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

Epstein, D. (2012). *The duality of information policy debates: The case of the Internet Governance Forum* [Cornell University]. https://ecommons.cornell.edu/items/9a225f60-6882-4e6e-98d4-d44c0eead532

Ermoshina, K., & Musiani, F. (2022). *Concealing for freedom: The making of encryption, secure messaging and digital liberties*. Mattering Press.

Ermoshina, K., & Musiani, F. (2025). Safer spaces by design? Federated socio-technical architectures in content moderation. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.1827

Ferret, N. (2025). The realm of digital content regulation as a social space: Sociogenesis of moderation norms and policies on Twitch platform. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.2004

Fishman, B. (2019). Crossroads: Counter-terrorism and the Internet. *Texas National Security Review*, *2*(2), 82–100.

Gillespie, T. (2017). Governance of and by platforms. In *SAGE handbook of social media* (pp. 254–278). SAGE Publications.

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gillespie, T., Aufderheide, P., Carmi, E., Gerrard, Y., Gorwa, R., Matamoros-Fernández, A., Roberts, S. T., Sinnreich, A., & West, S. M. (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

Gorwa, R. (2019). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review*, *8*(2), 1–22. https://doi.org/10.14763/2019.2.1407

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1). https://doi.org/10.1177/2053951719897945

Griffin, R. (2023a). From brand safety to suitability: Advertisers in platform governance. *Internet Policy Review*, *12*(3). https://doi.org/10.14763/2023.3.1716

Griffin, R. (2023b). 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

Grison, T., Julliard, V., Alié, F., & Ecrement, V. (2023). La modération abusive sur Twitter: Étude de cas sur l'invisibilisation des contenus LGBT et TDS en ligne [Abusive moderation on Twitter: A case study on the invisibility of LGBT and TDS content online]. *Réseaux*, *237*(1), 119–149. https://doi.org/10.3917/res.237.0119

Grover, R. (2022). The geopolitics of digital rights activism: Evaluating civil society's role in the promises of multistakeholder internet governance. *Telecommunications Policy*, *46*(10). https://doi.org/10.1016/j.telpol.2022.102437

Hallinan, B., Reynolds, C.J., Kuperberg, Y., & Rothenstein, O. (2025). Aspirational platform governance: How creators legitimise content moderation through accusations of bias. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.1829

Hill, S. (2025). Stop hate for profit: Evaluating the mobilisation of advertisers and the advertising industry to regulate content moderation on digital platforms. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.1825

Han, C., Seering, J., Kumar, D., Hancock, J. T., & Durumeric, Z. (2023). Hate raids on Twitch: Echoes of the past, new modalities, and implications for platform governance. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW1), 1–28. https://doi.org/10.1145/3579609

Haristya, S. (2020). The efficacy of civil society in global internet governance. *Internet Histories*, *4*(3), 252–270. https://doi.org/10.1080/24701475.2020.1769892

Helberger, N., Pierson, J., & Poell, T. (2017). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, *34*(1), 1–14. https://doi.org/10.1080/01972243.2017.1391913

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

Keck, M. E., & Sikkink, K. A. (2014). *Activists beyond borders: Advocacy networks in international politics*. Cornell University Press.

Keller, D. (2022). Lawful but awful? Control over legal speech by platforms, governments, and internet users. *The University of Chicago Law Review Online*, *1*, 1–21.

Kerkhof, J. (2025). Article 22 Digital Services Act: Building trust with trusted flaggers. *Internet Policy*

Review, *14*(1). https://doi.org/10.14763/2025.1.1828

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

Klonick, K. (2020). The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression. *The Yale Law Journal*, *129*(8), 2232–2605.

Kwak, J. (2014). Cultural capture and the financial crisis. In D. Carpenter & D. A. Moss (Eds.), *Preventing regulatory capture: Special interest influence and how to limit it* (pp. 71–98). Cambridge University Press.

La Quadrature du Net. (2020). *Loi Avia: Nos observations devant le Conseil Constitutionnel [Avia Act: Our observations before the Constitutional Council]*. La Quadrature du Net. https://www.laquadrature.net/2020/05/26/loi-avia-nos-observations-devant-le-conseil-constitutionnel/

Loveluck, B. (2016). Le vigilantisme numérique, entre dénonciation et sanction: Auto-justice en ligne et agencements de la visibilité [Digital vigilantism, between denunciation and sanction: Online self-justice and arrangements of visibility]. *Politix*, *115*(3), 127–153. https://doi.org/10.3917/pox.115.0127

Mansoux, A., & Roscam Abbing, R. (2020). Seven theses on the fediverse and the becoming of FLOSS. In K. Gansing & Luchs, I. (Eds.), *The eternal network: The ends and becomings of network culture* (pp. 124–140). Institute for Network Cultures and Transmediale.

Mărcuț, M. (2017). Crystallizing the EU digital policy. In M. Mărcuț (Ed.), *Crystalizing the EU digital policy: An exploration into the digital single market* (pp. 109–176). Springer International Publishing.

Matias, J., Johnson, A., Boesel, W., Keegan, B., Friedman, J., & DeTar, C. (2015). *Reporting, reviewing, and responding to harassment on Twitter*. Women, action, and the media (WAM!). https://ssrn.com/abstract=2602018

McSherry, C., York, J., & Cohn, C. (2018). *Private censorship is not the best way to fight hate or defend democracy: Here are some better ideas*. Electronic Frontier Foundation. https://www.eff.org/deeplinks/2018/01/private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some

Medzini, R. (2021). Credibility in enhanced self-regulation: The case of the European data protection regime. *Policy & Internet*, *13*(3), 383–406. https://doi.org/10.1002/poi3.251

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

Metz, J. (2015). *The European Commission, expert groups and the policy process: Demystifying technocratic governance*. Palgrave Macmillan.

Michalon, B. (2025). The role of civil society organisations in co-regulating online hate speech in the EU: a bounded empowerment. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.1826

Mouffe, C. (2010). Civil society, democratic values and human rights. In *Globality, democracy and civil society* (pp. 109–125). Routledge.

Mueller, M., Mathiason, J., & Klein, H. (2007). The internet and global governance: Principles and norms for a new regime. *Global Governance*, *13*, 237–254. https://doi.org/10.1163/19426720-01302007

Musiani, F., Cogburn, D. L., DeNardis, L., & Levinson, N. S. (Eds.). (2016). *The turn to infrastructure in internet governance*. Palgrave Macmillan.

Myers-West, S. (2017). Raging against the machine: Network gatekeeping and collective action on social media platforms. *Media and Communication*, *5*(3), 28–36. https://doi.org/10.17645/mac.v5i3.989

Nakamura, L. (2015). The unwanted labour of social media: Women of colour call-out culture as venture community management. *New Formations*, *86*, 106–112. https://doi.org/10.3898/NEWF.86.06.2015

Nielsen, R. K., & Ganter, S. A. (2022). *The power of platforms: Shaping media and society*. Oxford University Press.

Palladino, N., Redeker, D., & Celeste, E. (2025). Civil society's role in constitutionalising global content governance. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.1830

Paloque-Bergès, C. (2018). *Qu'est-ce qu'un forum internet? Une généalogie historique au prisme des cultures savantes numériques [What is an internet forum? A historical genealogy through the prism of digital scholarly cultures]*. OpenEdition Press.

Patry, W. (2009). *Moral panics and the copyright wars*. Oxford University Press.

Postigo, H. (2009). America online volunteers: Lessons from an early co-production community. *International Journal of Cultural Studies*, *12*(5), 451–469. https://doi.org/10.1177/1367877909337858

Radu, R. (2019). *Negotiating internet governance*. Oxford University Press.

Raymond, M., & DeNardis, L. (2015). Multistakeholderism: Anatomy of an inchoate global institution. *International Theory*, *7*(3), 572–616. https://doi.org/10.1017/S1752971915000081

Reynolds, C., & Hallinan, B. (2024). User-generated accountability: Public participation in algorithmic governance on YouTube. *New Media & Society*, *26*(9), 5107–5129. https://doi.org/10.1177/14614448241251791

Rochefort, A. (2025). Online advertising, content moderation, and corporate accountability: a civil society perspective. *Internet Policy Review*, *14*(1). https://policyreview.info/articles/news/online-advertising-content-moderation

Schirch, L. (2025). The case for prosocial tech design governance. *Internet Policy Review*, *14*(1). https://policyreview.info/articles/news/prosocial-tech-design-governance

Siapera, E., & Farries, E. (2025). Platform governance and civil society organisations: Tensions between reform and revolution continuum. Internet Policy Review, 14(1). https://doi.org/10.14763/2025.1.2002

Spencer-Smith, C., & Tales, T. (2025). Labour pains: Content moderation challenges in Mastodon growth. *Internet Policy Review*, *14*(1). https://doi.org/10.14763/2025.1.1831

Stigler, G. J. (2021). The theory of economic regulation. *The Bell Journal of Economics and Management Science*, *2*(1), 3–21. https://doi.org/10.2307/3003160

Sybert, J. (2022). The demise of #NSFW: Contested platform governance and Tumblr's 2018 adult content ban. *New Media & Society*, *24*(10), 2311–2331. https://doi.org/10.1177/1461444821996715

Tjahja, N., Meyer, T., & Shahin, J. (2021). What is civil society and who represents civil society at the

IGF? An analysis of civil society typologies in internet governance. *Telecommunications Policy*, *45*(6). https://doi.org/10.1016/j.telpol.2021.102141

Zuckerman, E. (2020). *The case for digital public infrastructure*. Knight First Amendment Institute at Columbia University. https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure

Zuckerman, E., & Rajendra-Nicolucci, C. (2023). From community governance to customer service and back again: Re-examining pre-web models of online governance to address platforms' crisis of legitimacy. *Social Media + Society*, *9*(3). https://doi.org/10.1177/20563051231196864

Published by

Alexander von Humboldt Institute for Internet and Society

CAIS Research for the Digital Age

in cooperation with

CREATe

cnrs centre internet et société

UOC R&I IN3 Internet interdisciplinary Institute Universitat Oberta de Catalunya

University of Tartu Johan Skytte Institute of Political Studies