INTERNET POLICY REVIEW
Journal on internet regulation

Volume 14 **Issue** 1

RESEARCH ARTICLE

OPEN ACCESS

PEER REVIEWED

Civil society's role in constitutionalising global content governance

Nicola Palladino *University of Salerno*
Dennis Redeker *University of Bremen*
Edoardo Celeste *Dublin City University*

DOI: https://doi.org/10.14763/2025.1.1830

Published: 31 March 2025
Received: 5 April 2024 **Accepted:** 22 October 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: 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

Keywords: Governance, Digital constitutionalism, Civil society, Network analysis

Abstract: This article examines global content governance on social media platforms through the lens of digital constitutionalism, which explores how fundamental rights can be embedded within the socio-technical architecture of digital technologies. It highlights the often-overlooked role of civil society in articulating digital rights and principles. In addition to performing a watchdog function and raising awareness about the human rights implications of digital technologies, we argue that civil society organisations play a constitutionalising role, acting as a bridge between international human rights law and platform governance. Above all, by engaging in global conversations, civil society organisations may facilitate the emergence and dissemination of a set of shared principles and rules. By conducting a semantic network analysis on 44 digital bills of rights that were drafted by civil society organisations and addressed content governance issues, the article aims to identify emerging principles as well as to study their alignment with human rights standards, their relationships, and evolution over time. The findings highlight how civil society initiatives have effectively led to a convergence of expectations around a common set of principles for content moderation, which could both pressure and support platforms and policymakers to strike a balance between freedom of expression and protecting people and democratic institutions from harm and disinformation.

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

The rise of the internet and social media came along with great expectations about their impact on democracy and public life. Especially after the so-called Arab Spring (Poell & Van Dijck, 2015), online platforms have been seen as "an infrastructure capable of revitalizing and extending the public sphere" (Santaniello et al., 2016). From this perspective, scholars have looked at some of their key features, such as their accessibility and affordability as well as the possibility to engage in interactive and horizontal flows of communication as a means to overcome traditional censorship, coordinate political movements, or foster deliberative democracy (Chadwick, 2009; Farrell, 2012). Social media have been compared to the Habermasian 'coffee shops' of eighteenth-century England, public fora capable of reversing the degradation of the public sphere due to the rise of traditional, one-to-many, mass media (Habermas, 1991), by granting unfiltered access to a plurality of sources of information and the possibility to have a say in public debates (Bimber et al., 2012; Bennett & Segerberg, 2012). However, recent events, such as the dissemination of disinformation campaigns during current conflicts (Okholm et al., 2024) or pivotal elections (Benkler et al., 2018; Bennett & Livingstone, 2020) have shown how social media could also easily resemble Munich's beer halls during the 1920s. In these environments, indeed, malicious actors can leverage people's vulnerabilities and frustration to manipulate public opinion and undermine democratic values and processes.

In the last few years, the awareness and willingness have matured among policymakers and stakeholders about the necessity to regulate content on social media platforms. However, identifying a proper set of norms for this purpose has proven to be extremely challenging, not only in light of the difficulty of reaching a consensus on which rules to adopt, but also due to the transnational nature of social media environments (Celeste et al., 2023). Social media platforms' structures do not align with national borders. They allow transborder communication flows between people located in different countries: what is published is potentially visible from all over the planet. Likewise, platforms' physical, economic and social infrastructure - from data centres to content moderation contractors - is distributed across the globe. States have proven to be able to effectively block the access to online content forbidden by their respective legal systems and to prosecute au-

thors and platforms within their territories (Tsagourias, 2021). Nevertheless, the quest for a universal set of rules for content governance remains a pressing need to address the interests and concerns of all stakeholders involved, due to various factors such as the overwhelming volume of content flooding social media every day, the possibility of circumventing national laws, platforms' ease in standardising their processes, and users' demand of clarity on permissible actions. Furthermore, if we rule out authoritarian regimes actively engaged in censorship practices, another crucial issue in the context of democratic countries is how to balance competing interests and values, or, in other terms, how to protect citizens and democratic institutions from harmful contents without violating fundamental rights and democratic values. This is a question to which different national and cultural traditions offer different answers, and that would require a certain degree of harmonisation.

Acknowledging the transnational nature of social media also entails recognising how they blur the boundaries between public and private spaces (Gillespie, 2018; Jørgensen & Zuleta, 2020). On the one hand, social media platforms represent private spheres governed by rules defined by owners and managers based on their commercial interests, without any legal obligations to consider user claims. On the other hand, it is widely acknowledged that "these private online spaces have acquired a public, if not 'constitutional', relevance" (Celeste et al., 2023, p. 11), given the increasing role they play in our social, economic, and political lives. As a consequence, when platforms intervene in content governance, they also perform delicate public functions that impact people's rights. Both violations of fundamental rights and the related necessary safeguards must navigate through architectures, internal policies, and organisational routines of social media platforms. For these reasons, it could be argued that the transnational nature of social media poses a primary obstacle to directly applying international human rights standards to content moderation, which may initially appear to be the most straightforward solution to achieving a shared global standard. Indeed, international human rights law, with its focus on nation states, lacks direct influence over private companies governing social media platforms. Moreover, its articulation of general principles seems inadequate for addressing the "complex socio-technical environment such as platform content moderation" (Celeste et al., 2023, p. 63).

This article aims to analyse the role that civil society can play in overcoming this limitation by contributing to the development of a *digital constitutionalism* framework tailored for social media platforms. The second section proposes an outline of the digital constitutionalism approach. Mainly based on societal constitutional-

ism theory and Science and Technology Studies, it highlights that constitutionalisation in the digital realm can be conceived as a hybrid process involving the generalisation and respecification of constitutional functions into specifically tailored socio-technical arrangements. In the third section, the role of civil society in the field of digital governance and in the digital constitutionalisation process is explored in detail. In addition to performing a watchdog function and raising awareness about the human rights implications of digital technologies, civil society organisations act as a bridge between international human rights law and platform governance. Above all and despite a series of inherent limitations of these actors, by engaging in global conversations, civil society organisations may discursively facilitate the emergence and dissemination of a shared set of principles and rules. Sections four and five engage into an empirical analysis of 44 digital bills of rights drafted by civil society organisations, which address content governance issues (please consult the full list in the appendix). By conducting a semantic network analysis on this textual corpus, the paper aims to highlight the relationships among the normative principles proposed by civil society organisations and their evolution over time. The concluding section shows how civil society initiates a process of 'translation' of international human rights principles into more granular norms that can be applied in platform operations. Civil society's digital bills of rights lead to a convergence of expectations around a common set of principles for content moderation, which might then be incorporated into legislative initiatives and private policies. Civil society efforts, while undoubtedly valuable, are far from sufficient to guarantee the implementation of a common set of content moderation rules based on international human rights law. However, civil society organisations are uniquely positioned to serve as intermediaries in this process of constitutionalisation, performing necessary 'bridging' and 'translation' functions—an essential preliminary step toward this goal.

2. Constitutionalising content governance: digital and societal constitutionalism

In a very first instance, we can define 'digital constitutionalism' as a "common term to connect a constellation of initiatives that have sought to articulate a set of political rights, governance norms, and limitations on the exercise of power on the Internet" (Redeker et al., 2018, p. 303). This definition focuses on digital constitutionalism as a movement, stemming from the political goals and will of a plurality of subjects concerned by the implications of digital technologies. In this regard, Celeste distinguishes between digital constitutionalism as "the ideology which aims to establish and to ensure the existence of a normative framework for the

protection of fundamental rights and the balancing of powers in the digital environment" (Celeste, 2019, p. 13) and digital constitutionalisation as the process of norms production for this purpose. Digital constitutionalism trace-back to the earliest stage of internet governance discussions, as "the first attempts to draft and formalize an Internet constitution and bill of rights [...] emerged soon after the establishment of the private Domain Name System regime" (Palladino and Santaniello, 2021, p. 52).

There are many ways to conceptualise the substantive content of digital constitutionalism and the practices of digital constitutionalisation (Celeste, 2019). One of the main sources of interpretation is represented by Gunter Teubner's societal constitutionalism theory (Teubner, 2004; Palladino, 2021, 2023) insofar as it deliberately works on a transnational scale and permits to address the transposition of constitutional principles into digital socio-technical architectures. The framework of societal constitutionalism draws upon the process of social differentiation described in Luhmann's (1975) social system theory. In this view, we can conceive society as articulated into a series of subsystems, each of them focused on a specific social function or sphere of human activity. As a social subsystem becomes more structured and autonomous, it gives rise to its own systemic logic based on a specific communicative media. The communicative media is the universal value giving sense to actions within the subsystems and then allowing actors to interact on a common ground. So, to give an example, the logic of the economic subsystem could be defined in terms of 'maximisation of profit', which in turn depends on the possibility to give a money-value to entities and interactions. The more a subsystem becomes relevant to the social system as a whole, the more likely it is to exhibit what Teubner (Teubner, 2011, 2012) describes as 'expansionist' and 'totalising' tendencies. These terms denote the inclination of the subsystem's logic to permeate other social spheres, to secure its reproduction at the expense of compromising the integrity and autonomy of individuals and communities.

It is worth recalling that societal constitutionalism adopts an inherently transnational perspective, well-suited to address the issues raised by social media. Indeed, while states may impose some constraints, social subsystems themselves do not conform to national borders. Therefore, using our example, economic operators from different countries tend to interact within a broader economic subsystem, speaking the common language of money and profit. In the last few decades, terms such as information society (Webster, 2002), digital society, network society (Castells, 2009) and platform society (Dijck et al., 2018) have emerged, suggesting that activities centered around the internet and digital technologies have become

increasingly recognisable as a distinct domain, while also becoming more relevant to society as a whole. It has evolved into an autonomous social subsystem, in which the logic is the digitalisation of society, conceived as the ongoing process of conversion of social reality into digital information in order to be further processed and elaborated to extract new information with added value (Palladino, 2023a). Its communicative media isthe 'code' (Lessig, 2009), a combination of software, hardware and operational routine that defines the architecture of cyberspace and enables actors' interaction.

As a part of the digitalisation process, content governance profoundly impacts the integrity and autonomy of individuals and communities in at least three main ways. First, by preventing access to and expression of opinions through architectural means, such as blocking and filtering. Second, content governance encompasses the transformation of communicative flows into engagement metrics, in order to prioritise and promote content that generates higher online traffic (Gillespie, 2018). Third, there is a systematic collection of user data to engineer microtargeting strategies (Zuboff, 2019). These practices may lead to automated censorship, systemic surveillance, and manipulation, further fueling well-known phenomena such as polarisation, filter bubbles, echo chambers, and the dissemination of hate speech and fake news.

In the perspective of societal constitutionalism, the constitutionalisation process occurs through a process of 'generalisation and respecification' of fundamental rights (Celeste, 2023). This involves abstracting the essential functions of fundamental rights and adapting them to the specific logic and communicative media of a subsystem. Consequently, it becomes conceivable to envision 'civic constitutions' beyond the nation-state paradigm, capable of addressing the challenges posed by the transnational nature of social media platforms. Fundamental rights, in this context, can be perceived as "social and legal counter-institutions" (Teubner, 2011, p. 210), countering the 'expansionist' and 'totalising' tendencies of digitalisation by fulfilling an 'inclusionary' and 'exclusionary' function. The latter roughly correspond to the empowering and limitative functions in classical constitutional thinking (Waldron, 2009). The inclusionary function grants individuals access to the communicative medium of the subsystem, referred to here as the 'code'. This encompasses the ability for people to scrutinise the workings of digital technologies and participate in their development. Conversely, the exclusionary function defines the boundaries of legitimate social media operations, preventing or sanctioning actions that jeopardise the autonomy and integrity of individuals and communities.

The basic assumption of the societal constitutionalism approach also implies three corollaries. First, "for fundamental rights to be truly effective in the social media environment, they must be translated and incorporated into their socio-technical architecture, including programming, algorithms, internal policies and operational routines" (Palladino, 2023a, p. 527). Secondly, as a process, digital constitutionalisation is articulated in different phases. At the beginning social norms that represent actors' expectations on platform functioning are created. Then, social norms are institutionalised into platforms' architecture and formalised into law. Finally, digital constitutionalisation is inherently a hybrid process (Santaniello et al., 2018). The adoption of limiting mechanisms by social media platforms is largely the result of external pressure exerted by civil society, governments, media, and other stakeholders. Moreover, the transposition of fundamental rights into platforms' architectures and operational routines is a complex task requiring the involvement of a wide range of actors with distinct roles and responsibilities. For example, states could ensure the bindingness of digital constitutionalism principles and norms; technical communities have the ability to translate them into operational standards and private companies to define the arrangements to concretely implement them into social media architectures (Palladino, 2023a; Celeste, 2019). The next section illustrates the pivotal role that civil society can play in all these aspects of the constitutionalisation process. Even without formalised authority, civil society is crucial in raising awareness about the implications of platforms functioning and rules, creating consensus around a consistent set of norms at the transnational level and favouring the translation of digital constitutionalism principles into operational arrangements.

3. The role of civil society in the process of digital constitutionalisation

Civil society is a broad term capturing a number of more or less organised forms of society beyond the state and the individual family (Chambers & Kopstein, 2006). Scholars at times disagree about the exact boundaries of what civil society entails but generally agree that its actors "focus predominantly on associational life rather than market or exchange relations" (Chambers & Kopstein, 2006, p. 363; see also Martens, 2002). Beyond the nation state, NGOs and other organised civil society actors act to increase the transparency and accountability of global governance by serving as watch dogs *vis-à-vis* global governance organisations. In the transnational field of digital governance, this includes both private and public organisations, such as social media platforms and the International Telecommunication Union, respectively. Organised civil society thus fills a gap in global governance as

citizens are only left with a 'notional accountability chain' between them and global organisations and processes (Scholte, 2004). Civil society, by being engaged with these organisations and processes, often advocates in favour of human and fundamental rights. Yet, in the field of global digital governance, its engagement traditionally goes beyond organised civil society and includes a variety of individuals, such as technical experts and 'politically engaged' academics (Van Eaten & Mueller, 2013). The reliance on engineers and academics, alongside private companies, can be connected to the long restraint of governments to involve themselves more prominently in a field characterised by a transnational scope and decentralised architectures, and the need to retain interoperability of technical standards (Mueller, 2010).

Activities of transnational civil society in the field of communication rights have a long tradition, as evidenced by the MacBridge Roundtables (1989-1998) or documents such as the 'People's Communication Charter' of 1999 (Padovani, 2005). While these roots and continuities should not be disregarded, the early 2000s are usually seen as a formative period of 'Internet governance', which - in extension - includes many of the questions faced in the governance of platforms and the possibility of human rights protection. Individuals, representatives of NGOs and other civil society groups joined the state-led 2003-2005 World Summit of the Information Society (WSIS), a three-year process that is now seen as the starting point of multistakeholder governance in the field (Palladino & Santaniello, 2020; Padovani & Tuzzi, 2004; Mueller et al., 2007), paving the way for the subsequent establishment of the Internet Governance Forum (IGF) and civil society involvement therein (Mueller, 2010). The WSIS has witnessed an emancipation of civil society in digital governance, e.g. through the drafting of a 'Civil Society Declaration to the World Summit on the Information Society' separate from the official declaration, and impacts beyond outcome documents in the form of "a contribution in broadening the agenda, a fruitful convergence of different civil society actors, and a continuity of interactions" (Padovani & Tuzzi, 2006, p. 66). Consequently, the Tunis Agenda, the official WSIS outcome document, called for "the full involvement of governments, the private sector, civil society and international organizations" (World Summit on the Information Society, 2005, Art. 26). As Raboy (2004) puts it, "the WSIS experience has transformed [the global governance as multistakeholder] framework most notably by sanctifying the place of global civil society as an organized force in this process" (p. 346). The subsequent establishment of the IGF as a project under the United Nations was relatively uncontroversial, primarily because it was "designed as a multi-stakeholder forum without any decision-making capacity" (Kleinwächter, 2007, p. 61).

Notably, in digital governance, while civil society actors may also act individually, initiatives to support a rights-based governance of digital technologies often emerged from networked civil society, in coalitions or so-called transnational advocacy networks (TANs), being already active in the context of the WSIS and continuing working into fora and organisations, such as the IGF, ICANN and NETmundial (Cogburn, 2017). Keck and Sikkink define TANs as to include "those actors working internationally on an issue, who are bound together by shared values, a common discourse, and dense exchanges of information" (Keck & Sikkink, 1998, p. 1).

The Communication Rights in the Information Society (CRIS) campaign was arguably the dominant TAN at the WSIS (Mueller et al., 2007; Franklin 2010). Spearheaded by the Association for Progressive Communications (APC), the CRIS campaign was successful in bringing into the conversation otherwise overlooked topics. It served to mobilise civil society, centralise civil society's efforts during the WSIS and to strengthen civil society actors' agenda-setting potential (Cogburn, 2017; Dany, 2012; Mueller et al., 2007). Franklin (2010) holds that the CRIS campaign successfully challenged dominant narratives that prioritise commercial and state interests over the human rights of individuals and communities. Digital bills of rights are key instruments to build community and mobilise civil society based on common advocacy positions. During the WSIS, members of the CRIS campaign successfully influenced the content of the Civil Society Declaration (Thomas, 2006), an early digital bill of rights, that represents an important reference point for the aimed entrenchment of rights and principles into digital governance debates.

In spite of its high visibility, the actual effectiveness of civil society in digital governance has been called into question by a number of scholars. While the WSIS and the IGF are not the only places to coordinate global content governance, they serve here to illustrate the challenges and the opportunities for civil society actors. Dany (2012) holds that, while civil society had ample opportunity to engage at the WSIS, structural power mechanisms limit the influence of participating NGOs, including resource and capacity gaps, the existence of informal processes that were out of reach for civil society and states' agenda-setting control. While raising awareness, building networks and supporting the institutionalisation of multistakeholder governance, civil society groups had limited ability to shape substantive policy decisions, as major outcomes were often driven by states and corporate actors (Mueller et al., 2007). In general, scholars have called into question the effectiveness of multistakeholder governance and the involvement of civil so-

ciety, referring to it as 'fiction' (Hofmann, 2016) or at least 'idealized' (Epstein, 2011). Carr (2015) further argues in this context that civil society "suffers from legitimacy challenges due to the difficulty of representing wide and diverse views and also due to the lack of oversight and accountability within civil society itself" (p. 657).

In the field of internet governance, TANs have been founded around more general issues such as 'communication rights' (CRIS campaign) to make non-state, non-corporate perspectives and demands more visible, or around more specific purposes such as to "imagine a feminist Internet" (Redeker, 2018, p. 5). They tend to be initiated or extended by larger NGOs involved but are usually open to other civil society actors. As argued above, a central activity of civil society, and TANs specifically, is to adopt advocacy texts that outline how key rights and principles ought to be respected by governments and private companies. These digital bills of rights then contribute to the constitutionalisation process of digital technologies at large. Celeste et al. (2024) argue that the African Declaration of Internet Rights and Freedoms and the Feminist Principles of the Internet serve as key advocacy tools in two ways: first, they aim to influence global and local discussions, raising awareness of critical issues and expanding the networks of support for the values they promote. Second, they actively engage with legislative and judicial institutions to shape people's rights regarding digital technologies through formal, institutional channels, including on the local and national level. These practices translate from general digital governance and human rights topics to more specific questions of content governance.

In practical terms, TANs, like civil society actors in general, analyse and closely observe issues related to corporate actions, such as content governance of platforms but also the regulatory (in)actions of states. Individual actors and networks of actors carry out a watchdog function, reporting human rights violations by both governments and companies, and advocating for common users, vulnerable groups and minorities, who may otherwise be marginalised or silenced in digital discourse. The pressure asserted through observation, reporting and scandalisation, in combination with an articulation of a positive vision on rights and principles can catalyse the process of generalisation and respecification of fundamental rights for the social media environment, ensuring that rights are not only recognised, but effectively protected in the online sphere. Hereby, civil society organisations have their work cut out to serve as intermediaries between abstract constitutional thinking and the practical governance of social media platforms. Through initiatives such as model policies, best practices guidelines, and multi-stakeholder

dialogues, civil society organisations provide actionable recommendations and granular norms for platforms to uphold human rights standards, while balancing competing interests such as freedom of expression and protection from harm. They engage on a global level, for instance with platforms, the UN or other organisations, but also in their national and regional constituencies.

Through civil society actors' engagement in TANs and coalitions for advocacy and research purposes, they foster the convergence of a set of normative values on how social media platforms ought to be governed across different jurisdictions. By sharing insights, exchanging best practices, and building coalitions, civil society can facilitate the development of a global standard for digital governance. This global standard might not only influence national legislation, but also shape the practices of multinational corporations operating in diverse regulatory environments. By advocating for consistent and rights-respecting approaches to content moderation more specifically, civil society organisations may contribute to the harmonisation of digital rights protection on a global scale.

4. Data and methods

To explore the role of civil society in the constitutionalisation of online content governance, we conducted a comprehensive examination of a collection of digital bills of rights sampled from the Digital Bills of Rights 1.0 data set (Redeker et al., 2024).[1]This data set, a collaborative initiative of researchers affiliated with the Digital Constitutionalism Network, offers an extensive repository of over 321 documents spanning from 1996 to the present day and engaging in digital constitutionalism initiatives. These documents encompass declarations of digital rights, resolutions, reports, and policy briefs, and represent contributions from diverse actors, including civil society organisations, governmental bodies, international entities, companies, and multistakeholder initiatives.

To date, it represents one of the most comprehensive collections of documents that entail normative demands, either aspirational or standard-setting, in the form of rights and principles in the context of the digital society. All the documents have been hand-coded for 19 meta codes (stakeholder group of the author/s, years of release, etc.) and 69 substantial content codes (privacy, transparency, etc.). From the Digital Constitutionalism Network Database we gathered a corpus of 44 documents, according to the following criteria: i) being authored by civil society enti-

1. A fully-searchable version of this data set can be found on the website of the Digital Constitutionalism Network, see https://digitalconstitutionalism.org/database

ties; ii) addressing the governance of online content, meant as the set of governing decisions regarding the hosting, dissemination, and presentation of user-generated content by internet service providers and online intermediaries. Selected documents have undergone qualitative coding (Saldaña, 2013) through the NVivo software (Mayring, 2019; Kaefer et al., 2015). In the first stage, principles were coded to closely reflect their original wording in the text. In the following stage, synonymous items were merged, and principles with fewer than three occurrences were grouped into broader categories, when appropriate. Coding procedures were collaboratively conducted by a team of three researchers autonomously coding portions of the data set. The validation process included inter-coder reliability checks, peer debriefing sessions, and member-checking techniques. Then, the coding results obtained through NVivo have been transposed into two matrices. The first one has documents in rows and normative items in columns. The second one is a squared matrix reporting the frequency at which two normative items co-occurred within the same paragraphs. Both of them have been analysed by resorting to the software Gephi (Bastian et al., 2009), in order to perform a semantic network analysis.

Semantic network analysis is a particular kind of social network analysis, which is focused on semantic elements in a text (or a corpus of texts), such as words, concepts, or themes. In this kind of network, words or concepts are the nodes connected to each other by 'ties' representing their co-occurrence in the same text or portion of text. The basic assumption here is that the meaning of a text, and, in truth, of the semantic elements themselves, emerges from the way in which semantic elements are interrelated and the structures they give rise (Segev, 2021).

Several social network analysis studies attempt to identify patterns in the strength and distribution of the ties between nodes (Granovetter, 1977), or consider the position of nodes within the network, and densely connected clusters (Borgatti et al., 2013). In our case, the analysis has been performed to reveal different kinds of relationships between charters and principles and lead to data-driven classification of items. In more detail, the analysis has been performed to explore to what extent civil society initiatives gave rise to a consistent normative framework, the articulation of the principles they proposed and their alignment with digital constitutionalism assumptions. The coding activity allowed us to identify more than 100 different normative elements. After aggregating semantically overlapping items and excluding the ones with a lower frequency, this number has been reduced to 62 items, which have been employed for the semantic network analysis.

5. Civil society discourse on content moderation

Figure 1 represents a bipartite network connecting documents and normative items, and it is based on a matrix reporting if a particular normative item is mentioned in a document (1 mentioned; 0 not mentioned). It confirms that civil society's discourses on content moderation focus on a relatively limited number of principles. But, probably, the most relevant indication we can draw from this graph is that all these initiatives give rise to a unique network without any separated components. The network shows the characteristics of so-called 'small world' networks (namely low graph density, small network diameter, short average path lengths). In this case, it indicates that, although actors could focus on different sets of normative items, their discourses tend to overlap to some degree. There are no isolated or loosely connected groups of items, meaning that there are no subgroups proposing radically alternative views on content moderation.

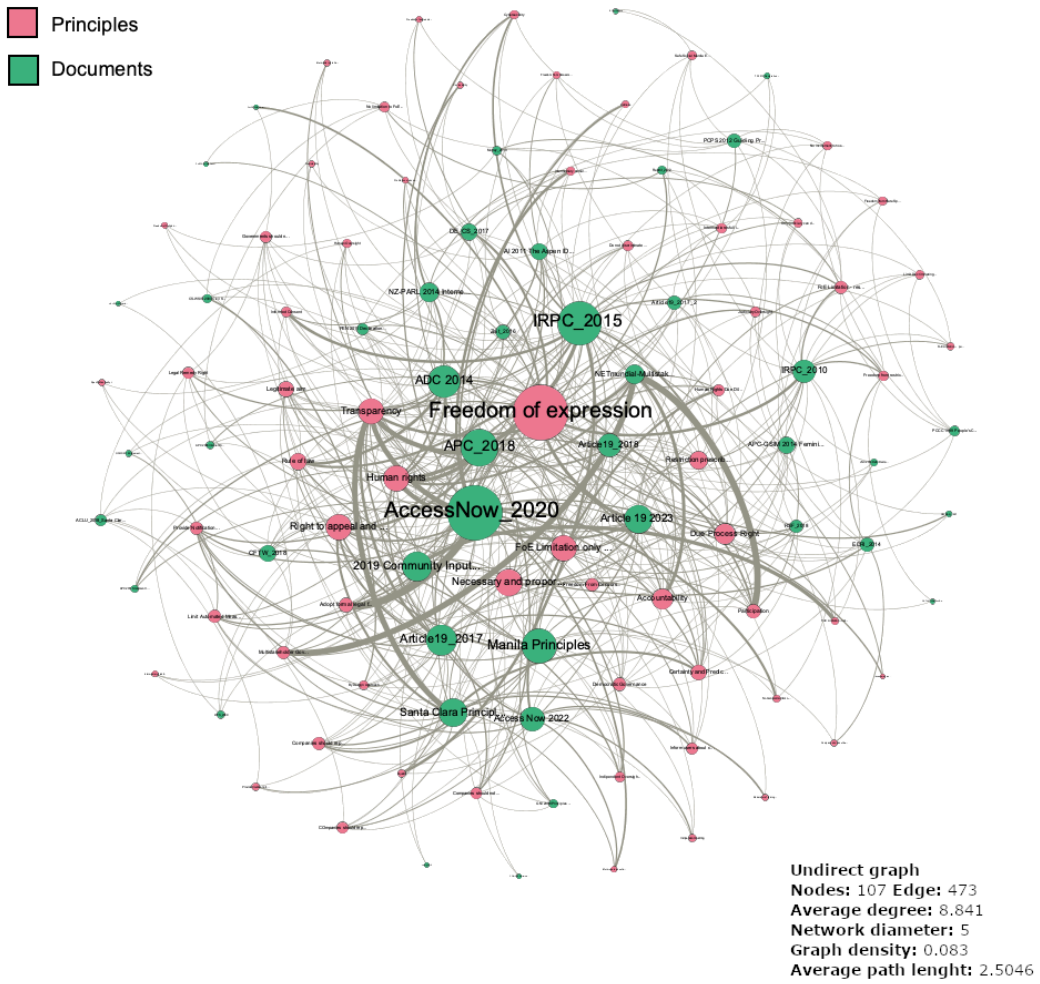

FIGURE 1: Documents and principles network

TABLE 1: Most frequent normative items

NORMATIVE ITEM	DOCUMENTS	REFERENCES
Freedom of expression	41	83
Necessary and proportionate	19	29
Human rights	18	59
Transparency	18	70
Right to appeal and remedy	18	36
FoE limitation only in accordance with human right standards	16	29
Accountability	14	25

NORMATIVE ITEM	DOCUMENTS	REFERENCES
Due process right	14	17
Restriction prescribed by law	12	14
Rule of law	11	23

Table 1 reports the most-cited principles in the corpus. Almost all the digital bills of rights refer to freedom of expression, which figures as the main concern of civil society organisations when dealing with content issues. This result reflects a long-standing tradition within communication fields, which recognises freedom of expression as a cornerstone of human rights protection and has also shaped the approach of civil society's initiatives on digital rights (Gill et al., 2015; Kuleska, 2008). Not by chance, freedom of expression is placed very often at the heart of a human rights and democratic value system, as the Principles on Freedom of Expression and Privacy drafted by the Global Network Initiative testify to:

> *Freedom of opinion and expression is a human right and guarantor of human dignity. The right to freedom of opinion and expression includes the freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers. Freedom of opinion and expression supports an informed citizenry and is vital to ensuring public and private sector accountability. Broad public access to information and the freedom to create and communicate ideas are critical to the advancement of knowledge, economic opportunity and human potential* (GNI, 2018, p.3).

The centrality of freedom of expression in civil society digital rights charters is also confirmed by the network analysis represented in Figure 1 and Table 2. The size of each node corresponds to its betweenness centrality, which refers to the node's capacity to connect with other nodes of the network. Even under these metrics, freedom of expression figures as the most relevant item. Moreover, the network analysis points out the existence of a backbone of recurring and interrelated principles in civil society's discourses on digital constitutionalism and content governance, which appears to be strictly rooted in the language and practices of international human rights law.

TABLE 2: Most relevant nodes' centrality values

PRINCIPLES	DEGREE CENTRALITY	BETWEENNESS CENTRALITY	DOCUMENTS	DEGREE CENTRALITY	BETWEENNESS CENTRALITY
Freedom of expression	41	1482.59	AccessNow 2020 Recommendations-On-Content-Governance-digital	41	913.34
Necessary and proportionate	19	216.10	IRPC 2015 The Charter of Human Rights and Principles for the Internet	32	734.79
Human rights	18	155.77	APC 2018 Content Regulation in the Digital Age	27	344.04
Transparency	18	161.88	The Manila Principles	25	313.50
Right to appeal and remedy	18	186.79	ADC 2014 African Declaration on Internet Rights and Freedoms	23	260.48
FoE limitation only in accordance with human right standards	18	187.15	Article 19 2017 Getting connected: Freedom of expression, telcos and ISPs	22	196.90
Accountability	14	124.94	NZ-CS 2019 Community Input on Christchurch Call	21	194.22
Due process right	14	117.17	Article 19 2023 Content Moderation Handbook	20	219.15
Restriction prescribed by law	12	88.12	The Santa Clara Principles 2.0	20	200.78
Rule of law	11	70.93	Access Now 2022 Declaration of principles for content and platform governance in times of crisis	17	94.20

6. From rights to procedures: generalisation and respecification of digital constitutionalism principles for social media platforms

Despite our corpus giving rise to a single network, it is possible to identify sub-groups of nodes strictly interconnected to observe how civil society has articulated a digital constitutionalism discourse on content governance. Figure 2 represents the graph of the normative elements coded in the corpus, as obtained from a squared co-occurrence matrix reporting how many times two normative items are

mentioned in the same paragraph. In doing so we can better understand how different concepts are related in civil society organisations' discourses. Indeed, the different colours partition the network into different communities of densely connected nodes, according to the Louvain 'modularity' algorithm (Blondel et al. 2008). The algorithm identifies groups of nodes that are highly connected internally and scarcely connected to each other (Borgatti et al., 2013). The main finding here is that actors' discourses on content governance seem not to overlap randomly but rather give rise to meaningful structure, articulated in four main clusters:

1) Human rights framework

The most important cluster collects a series of principles, implicitly or explicitly drawn from international human rights law, which appears to outline a 'Human rights framework' for content governance. Here we can witness a first attempt by civil society to apply well-established human rights principles to a novel reality, such as online content governance. For the most part it could be considered as an effort to define the exclusionary function of fundamental rights by setting the boundaries of legitimate content moderation operations.

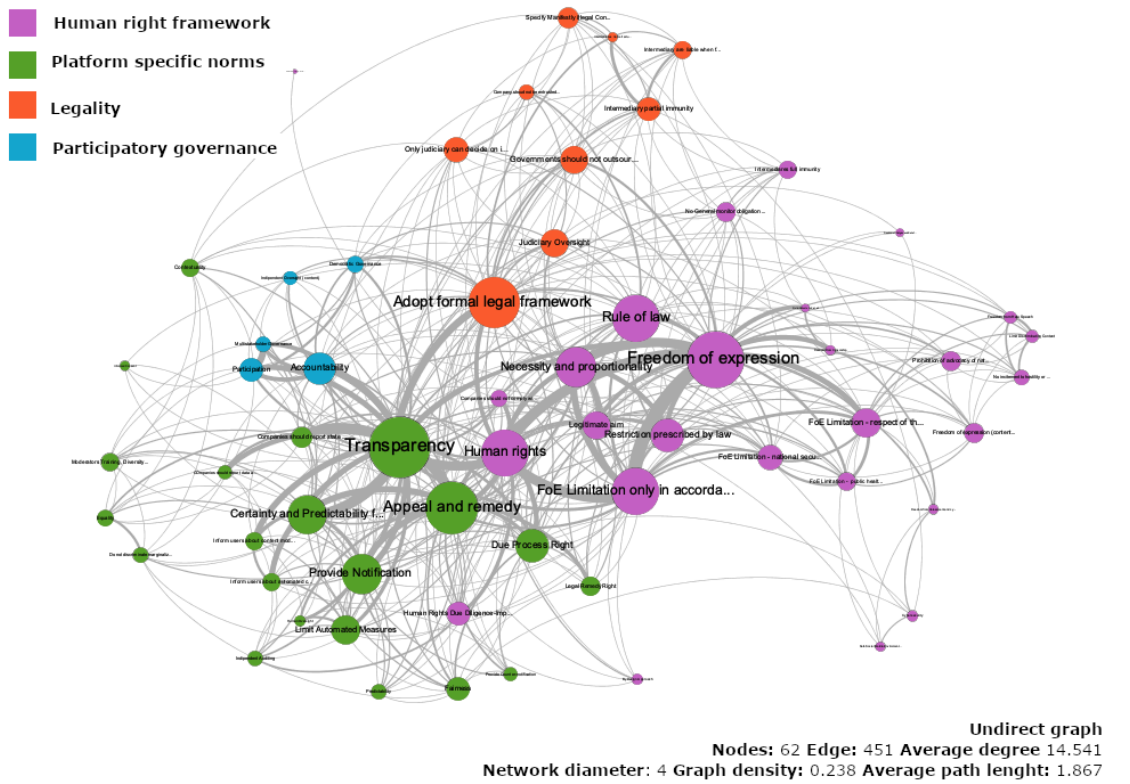

FIGURE 2: Semantic network

More than half of the corpus (26 documents) poses a generic, yet explicit, obligation to comply with international human rights standards and the rule of law. It is worth noting that this type of recommendation has typically been directed by civil society organisations towards states, referencing various international instruments, such as the Universal Declaration of Human Rights (UDHR) or the International Covenant on Civil and Political Rights (ICCPR). In this case, they are often extended to private companies as well in accordance with the United Nations Guiding Principles on Business and Human Rights (UNGPs), an instrument implementing the United Nations' 'Protect, Respect, and Remedy' framework.

Not surprisingly, freedom of expression plays a pivotal role in this cluster. It could be said it constitutes the 'frame' through which content moderation is interpreted and structured, meaning that the other rights and principles, for the vast majority, set the boundaries of freedom of expression, specifying its substantive content as well as the cases and the procedures according to which it could be legitimately limited. In the earliest documents of our data set, freedom of expression tends to be affirmed in quite absolute terms, with a priority over all other concerns. Consistently, these documents also tend to provide intermediaries with full immunity for third-party content. As the negative impact of illegal and harmful content emerges, civil society arguments become more sophisticated, incorporating competing concerns and rights, while still maintaining a strong foundation in international human rights law and the primacy of freedom of expression.

TABLE 3: Clusters and principles

CLUSTER 1 HUMAN RIGHTS FRAMEWORK	*CLUSTER 2 LEGALITY*	*CLUSTER 3 PLATFORM SPECIFIC NORMS*	*CLUSTER 4 PARTICIPATORY GOVERNANCE*
Freedom of expression *FoE limitation only in accordance with human right standards* *Human rights* *Rule of law* *Necessary and proportionate* *Restriction prescribed by law* *FoE limitation - respect of the rights or reputations, prevent harm* *Legitimate aim* *FoE limitation - national security or of public order* *Human rights due diligence-impact assessment* *No general monitor obligation* *Freedom of expression (content) limitation only when the rights and dignity of others are violated* *Prohibition of advocacy of national, racial or religious hatred* *Intermediaries full immunity*	*Adopt formal legal framework* *Governments should not outsource content regulation or enforcement to companies.* *Judiciary oversight* *Only judiciary can decide on illegal content* *Intermediary partial immunity* *Specify manifestly illegal content* *Intermediary are liable when fail to comply with adjudicatory order* *Company should not be entrusted with decision on the legality of content* *Intermediaries liable if actual knowledge*	*Transparency* *Right to appeal and remedy* *Provide notification to content providers-users* *Certainty and predictability for platform* *Due process right* *Limit automated measures* *Fairness* *Companies should report state requests* *Legal remedy right* *Inform users about automated content moderation* *Inform users about content moderation curation practices*	*Accountability* *Participation* *Democratic governance* *Multistakeholder governance* *Independent oversight*

CLUSTER 1 HUMAN RIGHTS FRAMEWORK	*CLUSTER 2 LEGALITY*	*CLUSTER 3 PLATFORM SPECIFIC NORMS*	*CLUSTER 4 PARTICIPATORY GOVERNANCE*
FoE limitation - public health or morals *Companies should not comply with State requests that are inconsistent with international human rights standards* *No incitement to hostility or violence* *Freedom from hate speech* *Limit discriminating content* *By design approach* *Cybersecurity* *Safe social media environment for vulnerable groups* *Freedom from censorship* *Freedom from harassment and cyberbullying* *Freedom of religion and belief on the Internet* *No limitation to FoE at all* *Intermediary liability - Do not involve infrastructure layer*		*Contextuality* *Moderators training, diversity, support* *Equality* *Do not discriminate marginalized groups content* *Independent auditing* *Predictability* *Companies should report data about content moderation activities* *Provide counter-notification* *Human oversight* *Informed consent*	

In the end, roughly half of the documents we analysed (21 cases) refer to the principles of the so-called Three-Part Test under article 19(3) of the ICCPR, according to which, admissible restrictions to freedom of expression must be:

> a) prescribed by law, in a clear and accessible manner, being formulated with sufficient precision to enable individuals to regulate their conduct accordingly;
>
> b) in pursuit of a legitimate aim, such as the 'respect of the rights or reputation of others' and the 'protection of national security, public order, or public health or morals'; or, according to Article 20 of the ICCPR, 'prohibiting propaganda for war and any advocacy of national, racial, or religious hatred' that constitutes incitement to discrimination, hostility, or violence.
>
> c) necessary and proportionate: there should be a direct and immediate connection between the expression and the identified threat, and the least restrictive measure capable of achieving a given legitimate objective should be imposed.

It is worth noting that civil society appears more inclined to acknowledge a legitimate exception to freedom of expression, when the latter is functional to prevent the endangerment of the autonomy and integrity of individuals and community, welcoming principles such as 'freedom from harassment and cyberbullying', 'limitation of discriminating contents' and 'safe social media enjoinment for vulnerable groups'. On the contrary, civil society is less inclined to accept limitations involving more abstract and collective values that could be employed to carry out undue

censorship. Issues such as 'hate speech' and 'disinformation' have been proven to be more burdensome and controversial. While we have some cases in our database proposing 'freedom from hate speech' and 'contrast to disinformation and fake news' as legitimate to be pursued in the social media environment, other actors criticised the notion of both 'hate speech' and 'disinformation' for their vagueness and lack of a definition grounded in international human rights law, which can ultimately lead to over-removal of lawful contents.

This human right framework is completed by further general prescriptions addressed to social media platform, according to which:

> i) 'companies should not comply with state requests that are inconsistent with international human rights standards';
>
> ii) 'states should not impose to intermediaries a general monitoring obligation'; resulting in a constant and indiscriminate screening of users content;
>
> iii) 'content moderation should not involve the infrastructural level', rendering entire websites and services inaccessible;
>
> iv) platforms 'should conduct human right due diligence and human right impact assessment' on a regular basis in order to identify, prevent, mitigate and account how their activities affect human rights.

2) Legality

Another set of principles revolves around establishing a legal framework for content moderation. This cluster further elaborates on the notion that limitations on freedom of expression should be governed by the rule of law. Taken as a whole, this group of norms outlines a framework wherein the law should not only strictly codify cases justifying action on content and accounts, but it must also "contain the definition of associated procedures—such as notice-and-action—and set high transparency standards for both states and online platforms. Most importantly, the legal framework must reinforce a clear distinction between the obligations of states and the responsibilities of private actors to protect users' human rights" (AccessNow, 2020, p. 26). In this regard, a crucial aspect of this cluster is the principle stating that only the judiciary or an independent authority can make decisions regarding the legality of third-party content. This principle is closely intertwined with others that define a precise stance on intermediary liability. In this view, governments should not delegate online content regulation and enforcement to platforms, and platforms should not be entrusted with decisions about illegal content.

Intermediaries could then be considered liable only when they fail to act upon acquiring actual knowledge about illegal content on their platforms, primarily through a court order, or when "manifestly illegal content" has been notified by a private entity, provided that what constitutes "manifestly illegal content" has been clearly determined by a legal framework.

3) Platform-specific norms

The third cluster collects a series of items elaborated by civil society in the last few years to move beyond a purely legal approach. These actors elaborated more granular norms which are specifically tailored for social media platforms to be implemented in their policy and socio-technical architecture in order to generalise and respecify the inclusionary and exclusionary functions of fundamental rights. Many civil society organisations have recognised that relying solely on the legal system cannot adequately address the issues posed by illegal and harmful content on social media due to the volume and speed of online communication. Legal actions, including strategic litigation, may be one important tool for civil society actors (Strobl, 2022; Celeste & Formici, 2024). Nonetheless, a court order may come too late, after content has already gone viral, causing irreversible damage, such as tarnishing the reputation of individuals, endangering vulnerable groups, or influencing electoral processes (Celeste 2021). Moreover, resorting to the judiciary for content removal requests or appeals may be prohibitively expensive and burdensome, resulting in many cases going unaddressed. Furthermore, defining illegal content is a challenging task, and even when defined, it may not be sufficient to prevent arbitrariness and unintended consequences without proper procedural guarantees. Finally, it cannot be ignored that platforms routinely engage in content curation and moderation practices, in addition to complying with legal regulations and state requests, to serve their own commercial interests and maximise user engagement on the platform (Gillespie, 2018).

Most civil society initiatives have aimed to establish procedural safeguards and uphold the principles of the rule of law within the privately regulated realm of social media. These efforts seek to mitigate the arbitrary exercise of power by platform operators. First, there is a call for platforms to ensure certainty and predictability by offering easily accessible terms of service and community standards. These guidelines should be openly accessible to the public, presented in the primary language of the user base, and communicated in straightforward language, avoiding complex technical or legal terminology (ARTICLE 19, 2017a; APC, 2018).

Platforms should also establish appeal and remedy procedures that do not prevent

users from resorting to traditional legal means. Instead, these procedures should offer a more direct path for users to contest supposedly unjustified decisions. According to the Santa Clara Principles (ACLU Foundation et al., 2018), appeal procedures include: "i) human review by a person or panel of persons that was not involved in the initial decision; ii) an opportunity to present additional information that will be considered in the review; iii) notification of the results of the review, and a statement of the reasoning sufficient to allow the user to understand the decision" (section 3). Companies are also requested to provide remedies, such as: "restoring eliminated content in case of an illegitimate or erroneous removal; providing a right to reply; with the same reach of the content that originated the complaint, offering an explanation of the measure; making information temporarily unavailable; providing notice to third parties; issuing apologies or corrections; providing economic compensation" (Access Now, 2020, p. 40). Platforms are also required to allocate sufficient financial and human resources to content moderation efforts. They must ensure that dedicated staff possess the necessary linguistic and cultural competencies to understand the context in which they operate and undergo proper training.

Most of the principles in this cluster relate to transparency. Besides making public content moderation and curation criteria, rules, and sanctions in their terms of service or community standards, platforms are asked to provide more detailed information about the means, methods, processes employed to carry out content moderation and curation practices. Social media platforms are required to regularly report and release data about their content removal activities broken down by country or region, if available, and category of rule violated encompassing: "total number of pieces of content actioned and accounts suspended; number of appeals of decisions to action content or suspend accounts; number of successful appeals that resulted in pieces of content or accounts being reinstated, and the number (or percentage) of unsuccessful appeals and; numbers reflecting enforcement of hate speech policies, by targeted group or characteristic [...] Number of discrete posts and accounts flagged, and number of discrete posts removed and accounts suspended, by source of flag (e.g., governments, trusted flaggers, users, different types of automated detection)" (Santa Clara Principles 2.0, 2021).

Furthermore, social media platforms are asked to provide each user whose content has been subject to moderation decisions, with a notification including information such as: "URL, content excerpt, and/or other information sufficient to allow identification of the content removed; the specific clause of the guidelines that the content was found to violate; how the content was detected and removed (flagged

by other users, governments, trusted flaggers, automated detection, or external legal or other complaint); explanation of the process through which the user can appeal the decision" (ACLU Foundation et al., 2018, n.p.).

Lastly, principles in this group also aim to establish some limits for the employment of automated systems for content moderation, such as hash-based tools or machine learning algorithms (Celeste et al., 2023). Some civil society organisations are skeptical about the possibility to carry out automated content removal in a rights-respecting way and call for the rejection of these methods (Global Forum for Media Development, 2019). The group of intellectuals convened by the Zeit Foundation, in its Charter of Digital Fundamental Rights of the European Union, specifies that "everyone has the right not to be the subject of computerised decisions which have significant consequences for their lives", and that "decisions which have ethical implications or which set a precedent may only be taken by a person" (2016, articles 7-8). Access Now is more likely to accept automated content moderation but "only in limited cases of manifestly illegal content that is not context-dependent, and should never be imposed as a legal obligation on platforms" (Access Now, 2020, p. 26).

In every instance, it is essential that individuals are informed when automated systems are employed for content monitoring, and they should have the opportunity to request a human review of such determinations. Additionally, companies must be obligated to elucidate how automated detection is utilised across various content categories, along with the rationale behind any content removal decisions. Overall, civil society organisations advocate for companies to address well-documented concerns regarding the accuracy, transparency, accountability, and fairness of automated content governance (Palladino, 2023). Automated systems should adhere to transparency standards, offering accessible explanations of their operations and decision-making criteria, as well as details on the processes involved, including avenues for appeal and redress.

4) Participatory governance

A last group of principles revolves around the idea of a participatory content governance, and it could be considered as a way to transpose the 'inclusionary function' of fundamental rights in the social media environment, for what concerns rule-making and decision-making processes. In this view, content governance should guarantee the meaningful participation of all affected parties to ensure that people have control over content moderation practices.

On the one hand content governance is supposed to be 'democratic', meaning that it requires a formal legal framework established by democratically elected bodies subject to public debate and scrutiny. On the other hand, platforms should: "Enable the meaningful participation of users in a timely manner and at different stages of the creation, implementation, and evaluation of content governance rules and technological developments; engage different groups of users, particularly those most affected by certain rules, decisions, or technologies; designate local points of contact to receive feedback and respond to users and civil society" (Access Now, 2020, article 10). This group also includes principles such as 'accountability' and 'independent oversight' (provided for example through recourse to an external entity, such as an oversight board or an industry-wide council), which should be conceived as further mechanisms to ensure people's control over content moderation practices and empower users.

Conclusions

The analysis conducted has shown how civil society digital constitutionalism initiatives have initiated a consistent and shared discourse on content moderation grounded on freedom of expression and other core human right law principles. While this discourse was originally "centered dominantly on protecting nascent spaces for online discourse against external coercion" (Bowel & Zittrain, 2020, p. 2), more recent documents acknowledge the necessity to define procedural safeguards for social media platforms' content moderation policies and limit their arbitrariness. While developing their discourse, civil society organisations did not deviate from the original international human rights law framework, but they further articulated it by engaging in a digital constitutionalisation effort. Civil society organisations produced norms generalising and respecifying international human rights law principles into the socio-technical architecture of social media platforms. Yet, this represents just the first stage of the process. These norms need to be further formalised into law, policies, technical standards and operational routine to become truly effective.

However, civil society-dominated TANs, by creating a common, consistent and articulated normative framework, can play a relevant role in defining a global standard or an internationally recognised set of norms for content moderation by putting pressure on states and companies and contributing to the draft of national and regional legislation and private policies. We already have some evidence that this process is ongoing. For example, the EU's Digital Services Act, which recently entered into force, includes some of the principles advanced by civil society, such

as i) transparency provisions requiring platforms to notify users about action on their content and account and their motivation, or regularly reporting about their content moderation activities; ii) the establishment of internal appeal and remedy procedures, which do not prevent resorting to ordinary or arbitral judicial mechanisms; iii) periodic human rights impact assessment. Similar principles have been also affirmed in the UNESCO Guidelines for the Governance of Digital Platforms (2023). In both cases, civil society organisations have contributed to the drafting of these documents taking part in consultation processes. Some of the civil society's principles discussed here have also been partially endorsed or implemented by major social media platforms (EFF, 2019).

Despite their natural limitations, civil society organisations continue to advocate for these principles and actively participate in the policy-making process. From the WSIS, where civil society entrenched its own role in internet governance, to the ongoing process of influencing platforms and state action, civil society actors - as part of a network or not - have articulated a clear vision on rights and principles. In this way, they hold the potential, in synergy with other stakeholders, to catalyse the establishment of a minimum global standard for content moderation, exerting pressure on both states and companies to embed fundamental rights in the socio-technical architecture of social media platforms.

References

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*. https://gephi.org/publications/gephi-bastian-feb09.pdf

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action. *Information Communication & Society*, *15*(5), 739–768. https://doi.org/10.1080/1369118X.2012.670661

Bennett, W., & Livingston, S. (2020). *The disinformation age*. Cambridge University Press.

Bimber, B., Flanagin, A., & Stohl, C. (2012). *Collective action in organizations: Interaction and engagement in an era of technological change*. Cambridge University Press.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10).

Bowers, J., & Zittrain, J. (2020). Answering impossible questions: Content governance in an age of disinformation. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-005

Carr, M. (2015). Power plays in global internet governance. *Millennium*, *43*(2), 640–659. https://doi.org/10.1177/0305829814562655

Castells, M. (2009). *The rise of the network society*. John Wiley & Sons.

Celeste, E. (2019). Digital constitutionalism: A new systematic theorisation. *International Review of Law, Computers & Technology*, *33*(1), 76–99. https://doi.org/10.1080/13600869.2019.1562604

Celeste, E. (2021). Digital punishment: Social media exclusion and the constitutionalising role of national courts. *International Review of Law, Computers & Technology*, *35*(2), 162–184. https://doi.org/10.1080/13600869.2021.1885106

Celeste, E. (2023). Internet bills of rights: Generalisation and re-specification towards a digital constitution. *Indiana Journal of Global Legal Studies*, *30*(2), 25–54.

Celeste, E., & Formici, G. (2024). Constitutionalizing mass surveillance in the EU: Civil society demands, judicial activism, and legislative inertia. *German Law Journal*, *25*(3), 427–446. https://doi.org/10.1017/glj.2023.105

Celeste, E., Iglesias Keller, C., & Redeker, D. (2024). Digital bills of rights. In G. De Gregorio, O. Pollicino, & P. Valcke (Eds.), *The Oxford handbook of digital constitutionalism* (1st ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198877820.013.17

Celeste, E., Palladino, N., Redeker, D., & Yilma, K. (2023). *The content governance dilemma: Digital constitutionalism, social media and the search for a global standard*. Springer International Publishing. https://link.springer.com/10.1007/978-3-031-32924-1

Chadwick, A. (2009). Web 2.0: New challenges for the study of e-democracy in an era of informational exuberance. *I/S: A Journal of Law and Policy for the Information Society*, *5*(1), 11–41.

Chambers, S., & Kopstein, J. (2006). Civil society and the state. In J. S. Dryzek, B. Honig, & A. Phillips (Eds.), *The Oxford handbook of political theory* (pp. 363–381). Oxford University Press.

Cogburn, D. L. (2017). *Transnational advocacy networks in the information society: Partners or pawns?* Springer.

Dany, C. (2012). *Global governance and NGO participation: Shaping the information society in the United Nations*. Routledge.

Dijck, J. van, Poell, T., & Waal, M. de. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Eeten, M. J., & Mueller, M. (2013). Where is the governance in Internet governance? *New Media & Society*, *15*(5), 720–736. https://doi.org/10.1177/1461444812462850

Epstein, D. (2011). Manufacturing internet policy language: The inner workings of the discourse construction at the Internet Governance Forum. *Research Conference on Communications, Information, and Internet Policy (TPRC) 39*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1989674

Farrell, H. (2012). The consequences of the internet for politics. *Annual Review of Political Science*, *15*(1), 35–52. https://doi.org/10.1146/annurev-polisci-030810-110815

Franklin, M. I. (2010). Digital dilemmas: Transnational politics in the twenty-first century. *The Brown Journal of World Affairs*, *16*(2), 67–85.

Gebhart, G., Crocker, A., Mackey, A., Opsahl, K., Tsukayama, H., Williams, J.L., & York, J.C. (2019). *Who has your back?* (Censorship). EFF. https://www.eff.org/wp/who-has-your-back-2019

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

Habermas, J. (1991). *The structural transformation of the public sphere*. Polity Press.

Hofmann, J. (2016). Multi-stakeholderism in internet governance: Putting a fiction into practice. *Journal of Cyber Policy*, *1*(1), 29–49. https://doi.org/10.1080/23738871.2016.1158303

Jørgensen, R. F., & Zuleta, L. (2020). Private governance of freedom of expression on social media platforms: EU content regulation through the lens of human rights standards. *Nordicom Review*, *41*(1), 51–67. https://doi.org/10.2478/nor-2020-0003

Keck, M. E., & Sikkink, K. (1998). *Activists beyond borders: Advocacy networks in international politics*. Cornell University Press.

Kleinwächter, W. (2007). The history of internet governance. In C. Möller & A. Amouroux (Eds.), *Governing the internet – freedom and regulation in the OSCE region* (pp. 41–66). OSCE.

Lessig, L. (2009). *Code 2.0*. Basic Books.

Martens, K. (2002). Mission impossible? Defining nongovernmental organizations. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *13*(3), 271–285.

Mueller, M. L. (2010). *Networks and states: The global politics of internet governance*. MIT Press.

Mueller, M. L., Kuerbis, B. N., & Pagé, C. (2007). Democratizing global communication? Global civil society and the campaign for communication rights in the information society. *International Journal of Communication*, *1*(1), 267–296.

Padovani, C. (2005). Debating communication imbalances from the MacBride Report to the World Summit on the Information Society: An analysis of a changing discourse. *Global Media and Communication*, *1*(3), 316–338. https://doi.org/10.1177/1742766505058127

Padovani, C., & Tuzzi, A. (2004). The WSIS as a world of words: Building a common vision of the information society? *Continuum*, *18*(3), 360–379. https://doi.org/10.1080/1030431042000256117

Padovani, C., & Tuzzi, A. (2006). Communication governance and the role of civil society: Reflections on participation and the changing scope of political action. In N. Carpentier & J. Servaes (Eds.), *Towards a sustainable information society: Deconstructing WSIS* (pp. 51–79). Intellect Books.

Palladino, N. (2021). Imbrigliare i giganti digitali nella rete del costituzionalismo ibrido. Spunti dall'approccio europeo alla governance dell'Intelligenza artificiale [Bridle big tech in hybrid constitutionalism: Insights from the European Union approach to artificial intelligence]. In M. Santaniello (Ed.), *Comunicazionepuntodoc*. Fausto Lupetti Editore.

Palladino, N. (2023a). A 'biased' emerging governance regime for artificial intelligence? How AI ethics get skewed moving from principles to practices. *Telecommunications Policy*, *47*(5), 102479. https://doi.org/10.1016/j.telpol.2022.102479

Palladino, N. (2023b). A digital constitutionalism framework for AI: Insight from the Artificial Intelligence Act. *Digital Politics*, *3*(3), 521–542.

Palladino, N., & Santaniello, M. (2020). *Legitimacy, power, and inequalities in the multistakeholder internet governance: Analyzing IANA transition*. Springer Nature.

Poell, T., & Dijck, J. (2015). Social media and activist communication. In C. Atton (Ed.), *The Routledge*

companion to alternative and community media (pp. 527–537). Routledge.

Raboy, M. (2004). The WSIS as a political space in global media governance. *Continuum: Journal of Media & Cultural Studies*, *18*(3), 345–359. https://doi.org/10.1080/1030431042000256108

Redeker, D. (2018). Exploring the bottom-up digital constitutionalism of the 'Feminist Principles of the Internet'. *Global Internet Governance Academic Network (GigaNet) Annual Symposium 2018*. http s://papers.ssrn.com/sol3/papers.cfm?abstract_id=3490214

Redeker, D., Gill, L., & Gasser, U. (2018). Towards digital constitutionalism? Mapping attempts to craft an Internet Bill of Rights. *International Communication Gazette*, *80*(4), 302–319. https://doi.org/10.1177/1748048518757121

Redeker, D., Palladino, N., Celeste, E., Santaniello, M., & Padovani, C. (2024). *Digital Bills of Rights* (Version 1.0.0) [Dataset]. GESIS Data Archive. https://doi.org/10.7802/2731

Santaniello, M., Blasio, E. D., Palladino, N., Selva, D., Nictolis, E. D., & Perna, S. (2016). Mapping the debate on internet constitution in the networked public sphere. *Comunicazione politica*, *3*, 327–354. https://doi.org/10.3270/84677

Santaniello, M., Palladino, N., Catone, M. C., & Diana, P. (2018). The language of digital constitutionalism and the role of national parliaments. *International Communication Gazette*, *80*(4), 320–336. https://doi.org/10.1177/1748048518757138

Santos Okholm, C., Ebrahimi Fard, A., & Ten Thij, M. (2024). Blocking the information war? Testing the effectiveness of the EU's censorship of Russian state propaganda among the fringe communities of Western Europe. *Internet Policy Review*, *13*(3). https://doi.org/10.14763/2024.3.1788

Scholte, J. A. (2004). Civil society and democratically accountable global governance. *Government and Opposition*, *39*(2), 211–233. https://doi.org/10.1111/j.1477-7053.2004.00121.x

Strobel, V. (2022). Strategic litigation and international internet law. In A. Golia Jr., M. C. Kettemann, & R. Kunz (Eds.), *Digital transformations in public international law* (pp. 261–284). Nomos.

Teubner, G. (2004). Societal constitutionalism: Alternatives to state-centred constitutional theory? In C. Joerges, I.-J. Sand, & G. Teubner (Eds.), *Transnational governance and constitutionalism*. Hart Publishing.

Teubner, G. (2012a). 5 Transnational fundamental rights: Horizontal effect. In *Constitutional fragments: Societal constitutionalism and globalization* (pp. 124–149). Oxford University Press. http s://doi.org/10.1093/acprof:oso/9780199644674.001.0001

Teubner, G. (2012b). *Constitutional fragments: Societal constitutionalism and globalization*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199644674.001.0001

Thomas, P. (2006). The Communication Rights in the Information Society (CRIS) campaign: Applying social movement theories to an analysis of global media reform. *International Communication Gazette*, *68*(4), 291–312. https://doi.org/10.1177/1748048506065763

Tsagourias, N. (2021). The legal status of cyberspace: Sovereignty redux? In N. Tsagourias & R. Buchan (Eds.), *Research handbook on international law and cyberspace*. Edward Elgar Publishing. http s://doi.org/10.4337/9781789904253.00010

Webster, F. (2002). *Theories of the information society*. Routledge.

World Summit on the Information Society. (2005). *Tunis agenda for the information society*. http://ww

w.itu.int/net/wsis/docs2/tunis/off/6rev1.html

Zuboff, S. (2019). *The age of surveillance capitalism*. Public Affairs.

Appendix: List of analysed documents

1. Access Now. (2020). *26 Recommendations on Content Governance*. https://www.accessnow.org/cms/assets/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf.
2. Access Now. (2022). *Declaration of principles for content and platform governance in times of crisis*. https://www.accessnow.org/publication/new-content-governance-in-crises-declaration/
3. ACLU Foundation et al. (2018). *The Santa Clara Principles on Transparency and Accountability in Content Moderation*. https://santaclaraprinciples.org.
4. ADC (African Declaration Coalition). (2014). *African Declaration on Internet Rights and Freedoms*. http://africaninternetrights.org/articles/.
5. APC (Association for Progressive Communications). (2011). *Statement Internet rights are human rights*. https://www.apc.org/en/project/internet-rights-are-human-rights
6. APC (Association for Progressive Communications). (2014). *Feminist Principles of the Internet*. https://feministinternet.org/sites/default/files/Feminist_principles_ of_the_internetv2-0.pdf.
7. APC (Association for Progressive Communications). (2018). *Content Regulation in the Digital Age*. https://www.ohchr.org/Documents/Issues/Opinion/ContentRegulation/APC.pdf.
8. Article 19. (2017a). *Getting connected: Freedom of expression, telcos and ISPs*. https://www.article19.org/wp-content/uploads/2017/06/ Final-Getting-Connected-2.pdf.
9. Article 19. (2017b). *Universal Declaration of Digital Rights*. https://www.article19.org/resources/internetofrights-creating-the-universaldeclaration-of-digital-rights/.
10. Article 19. (2018). *Self-regulation and 'hate speech' on social media platforms*. https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms_March2018.pdf
11. Article 19. (2023). *Content moderation and freedom of expression handbook*. https://www.article19.org/wp-content/uploads/2023/08/SM4P-Content-moderation-handbook-9-Aug-final.pdf
12. Aspen Institute. (2011). *The Aspen IDEA Principles*. http://www.umic.pt/images/stories/publicacoes5/Aspen%20IDEA%20Project.pdf. Association for Progressive Communications. 2006. Internet Rights Charter. https://www.apc.org/node/5677.
13. Cambodian Center for Independent Media. (2015). *Statement of Principles for Cambodian Internet Freedom*. http://www.cchrcambodia.org/media/files/press_release/573_201jsopfcife_en.pdf.

14. Chinese Academics. (2009). *China Internet Human Rights Declaration*. https://advox.globalvoices.org/2009/10/09/china-internet-human-rights-declaration/
15. Civil Society (WSIS). (2003). *Civil Society Declaration to the World Summit on the Information Society*. https://www.itu.int/net/wsis/docs/ geneva/civil-society-declaration.pdf.
16. Civil Society—TUAC. (2008). *Seoul Declaration*. https://thepublicvoice.org/PVevents/seoul08/seoul-declaration.pdf.
17. Contract for the Web. (2018). https://contractfortheweb.org.
18. DE-Civil Society. (2017). *Declaration on Freedom of Expression In response to the adoption of the Network Enforcement Law* ("Netzwerkdurchsetzungsgesetz") by the Federal Cabinet on April 5, 2017. https://deklaration-fuer-meinungsfreiheit.de/en/.
19. EDRi. (2014). *The Charter of Digital Rights*. https://edri.org/wpcontent/uploads/2014/06/EDRi_DigitalRightsCharter_web.pdf.
20. Gelman, R. G. (1997). *Declaration of Human Rights in Cyberspace*. www.be-in.com/10/rightsdec.html
21. GFMD (Global Forum for Media Development). (2019). *GFMD Statement on the Christchurch Call and Countering Violent Extremism Online*. https://drive.google.com/file/d/1N4EwiM7eITD6plQrYqJI02NayvCiMCT-/ view.
22. GNI (Global Network Initiative). (2008). *Principles on Freedom of Expression and Privacy*. https://www.globalnetworkinitiative.org/principles/ index.php.
23. iRights Coalition. (2015). *iRights*. http://irights.uk/the_irights/
24. IRPC (Internet Rights and Principles Coalition). (2010). *The Charter of Human Rights and Principles for the Internet*. http://internetrightsandprinciples.org/site/charter/.
25. IRPC (Internet Rights and Principles Coalition). (2014). *The Charter of Human Rights and Principles for the Internet 2.0 4th ed*. http://www.ohchr.org/Documents/Issues/Opinion/ Communications/InternetPrinciplesAndRightsCoalition.pdf.
26. Jarvis, Jeff. (2010). *A Bill of Rights in Cyberspace*. http://buzzmachine.com/2010/03/27/a-bill-of-rights-in-cyberspace/.
27. Just Net Coalition. (2014). *Delhi Declaration for a Just and Equitable Internet*. http://justnetcoalition.org/delhi-declaration.
28. Manila Principles. (2015). https://manilaprinciples.org/index.html
29. Murray, Jeffey. (2010). *A Bill of Rights for the Internet*. http://theitlawyer.blogspot.com/2010/10/bill-of-rights-for-internet.html.
30. NetMundial. (2014). *NetMundial Statement*. http://netmundial.br/wpcontent/uploads/2014/04/NETmundial- MultistakeholderDocument.pdf.
31. New Zealand Civil Society. (2019). *Community Input on Christchurch Call*. https://docs.google.com/document/d/10RadyVQUNu1H5D7x6IJVKbqmaeDeXre0Mk-FFNkIVxs/edit?ts=5cda8902#.
32. New Zealand Green Party. (2014). *Internet Rights and Freedoms Bill*. https://home.greens.org.nz/misc-documents/internet-rights-andfreedoms-

bill.
33. PCCC (People's Communication Charter Coalition). (1999). *People's Communication Charter*. http://www.pccharter.net/charteren.html.
34. Praxis Centre for Policy Studies. (2012). *Guiding Principles of Internet Freedom*. http://www.praxis.ee/fileadmin/tarmo/Projektid/ Valitsemine_ja_kodanike%C3%BChiskond/Praxis_Theses_Internet.pdf.
35. PEN International. (2011). *Declaration on Digital Freedom*. http://www.pen-international.org/declaration-on-digital-freedom-english/.
36. PHRF (Parliamentary Human Rights Foundation and Open Society Institute). (1997). *Open Internet Policy Principles*. http://mailman.anu.edu.au/pipermail/link/1997 March/026302.html.
37. Reddit. (2012). *The Digital Bill of Rights—A Crowd-sourced Declaration of Rights*. https://www.reddit. com/r/fia/comments/vuj37/ digital_bill_of_rights_1st_draft/.
38. Reporters Sans Frontiers. (2018). *International Declaration on Information and Democracy*. https://rsf.org/en/global-communication-andinformation-space-common-good-humankind. 140
39. Santa Clara Principles 2.0. (2021). https://santaclaraprinciples.org/
40. Tech Freedom et al. (2012). *Declaration of Internet Freedom*. http://declarationofinternetfreedom.org/.
41. UK Liberal Democrats. (2015). *Protecting your data online with a Digital Rights Bill. http://www.libdems.org.uk/protecting-your-data-onlinewith-a-digital-rights-bill*.
42. UNESCO. (2023). *Guidelines for the governance of digital platforms: Safeguarding freedom of expression and access to information through a multi-stakeholder approach*. https://unesdoc.unesco.org/ark:/48223/ pf0000387339
43. World Federation of Scientists. (2009). *Erice Declaration on Principles for Cyber Stability and Cyber Peace*. http://www.aps.org/units/fip/ newsletters/ 201109/barletta.cfm.
44. Zeit Foundation. (2016). *Charter of Digital Fundamental Rights of the European Union*. https://digitalcharta.eu/wp-content/uploads/ 2016/12/ Digital-Charta-EN.pdf.

Published by
Alexander von Humboldt Institute for Internet and Society

in cooperation with
CREATe
cnrs centre internet et societe
R&I IN3 Internet interdisciplinary Institute Universitat Oberta de Catalunya
University of Tartu Johan Skytte Institute of Political Studies